Tennessee Law Review

Volume 78
Issue 1 *2010-2011*

Article 1

2011

Volume 78 (2010-2011)

TENNESSEE LAW REVIEW

Member of the National Conference of Law Reviews

Volume 78 | Fall 2010 | Number 1

CONTENTS

ARTICLES

NONUSE AND EASEMENTS: CREATING A PLIABILITY REGIME OF PRIVATE EMINENT DOMAIN
Sally Brown Richardson 1

REDEEMING AN EMPTY PROMISE: PROCEDURAL JUSTICE, THE CRIME VICTIMS' RIGHTS ACT, AND THE VICTIM'S RIGHT TO BE REASONABLY PROTECTED FROM THE ACCUSED
Mary Margaret Giannini 47

LESS THAN PICTURE PERFECT: THE LEGAL RELATIONSHIP BETWEEN PHOTOGRAPHERS' RIGHTS AND LAW ENFORCEMENT
Morgan Leigh Manning 105

COMMENTS

COPYRIGHTING COUTURE: AN EXAMINATION OF FASHION DESIGN PROTECTION AND WHY THE DPPA AND IDPPPA ARE A STEP TOWARDS THE SOLUTION TO COUNTERFEIT CHIC
Sara R. Ellis 163

STRICTLY BUSINESS: A HISTORICAL NARRATIVE AND COMMENTARY ON ROCK AND ROLL BUSINESS PRACTICES
Jeff Carter 213

CASE NOTES

TORT LAW—COMPARATIVE FAULT—ORIGINAL TORTFEASOR RULE IN TENNESSEE
Dan Calvert 259

BANKRUPTCY LAW—DISCHARGEABILITY OF STUDENT LOAN DEBT—APPLICATION OF FEDERAL RULE OF CIVIL PROCEDURE 60(B)(4) TO DISCHARGE-BY-DECLARATION PROVISIONS
Dorothea K. Thompson 287

NONUSE AND EASEMENTS: CREATING A PLIABILITY REGIME OF PRIVATE EMINENT DOMAIN

SALLY BROWN RICHARDSON*

INTRODUCTION 2
PART I. THE CALABRESI AND MELAMED FRAMEWORK 5
PART II. THE INEFFICIENCY OF LOOKING THROUGH A PROPERTY LENS: THE AMERICAN COMMON LAW RULE AGAINST TERMINATION BY NONUSE 7
A. *The Common Law Rule Against Termination by Nonuse—a Classic Property Rule* 7
B. *The Inefficiency in Applying a Property Rule* 12
1. Inability to Predict the Future 12
2. Creation of Societal Loss 16
3. Encouraging Speculation and Division in Land 17
4. Giving Rise to Holdups 19
PART III. THE INEFFICIENCY OF LOOKING THROUGH A QUASI-PROPERTY LENS: THE CIVIL LAW RULE ALLOWING TERMINATION BY NONUSE 21
A. *The Civil Law Rule Allowing Termination by Nonuse, a Quasi-Property Rule* 22
B. *The Inefficiency in Applying a Quasi-Property Rule* 25
1. Inability to Predict Optimal Use 26
2. Allowing Underuse or Encouraging Overuse 27
PART IV. THE BELL AND PARCHOMOVSKY GLOSS 28
PART V. THE EFFICIENCY OF LOOKING THROUGH A PLIABILITY LENS: PRIVATE EMINENT DOMAIN BASED ON NONUSE 30
A. *Benefits of Applying Liability Rules to Nonused Easements* 31
1. Remedy Inefficiencies of the Common and Civil Law 31
2. Produce Superior Information 32
3. Assign Risk of Nonuse to Nonusing Party 33

---

* Associate at Skadden, Arps, Slate, Meagher & Flom. J.D./D.C.L., Louisiana State University; B.A., Georgetown University. This article was presented at Georgetown Law School in March 2010 at the first conference of the Association for Law, Property, and Society, and the author thanks all conference participants for their suggestions. The author also thanks James W. Bowers, Andrea B. Carroll, John M. Church, and Ronald J. Scalise Jr., for their comments on earlier drafts. All views and any errors herein are solely attributable to the author.

B. Creating a Pliability Regime: Private Eminent Domain Based on Nonuse 34
1. Setting Nonuse as the Trigger Point 35
a. Clearing the Initial Constitutional Hurdle 35
b. Determining the Point of Nonuse 37
2. Structuring the Compensation 41
CONCLUSION 44

INTRODUCTION

In their landmark work, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, Guido Calabresi and Douglas Melamed identified two distinct methods of protecting entitlements: liability rules and property standards.[1] Under *The Cathedral*'s construction, property standards protect the entitlement holder from the nonconsensual termination of the entitlement, while liability rules protect the non-holder and allow for the unilateral extinguishment of the entitlement.[2] After creating a conceptual framework of entitlement protections, *The Cathedral* developed a means of determining when each rule should be utilized.[3]

Real property rights almost unanimously have been subject to what *The Cathedral* considers to be property standards;[4] however, whether property standards are always the most efficient means of protecting all rights in real property is questionable. One particular area of real property that is in need of close inspection is the law of servitudes. Historically considered "an unspeakable quagmire," most of the law of servitudes has been carried forward for centuries, sometimes with little rhyme or reason.[5] During the final decades of the twentieth century, the drafters of the *Restatement (Third) of Property on Servitudes* recognized the discombobulated nature of

---

1. *See* Guido Calabresi & A. Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089, 1092 (1972) [hereinafter *The Cathedral*]. Calabresi and Melamed also discussed the concept of inalienable entitlements, i.e. those entitlements that cannot be transferred between a willing buyer and willing seller. *Id.* As this article discusses only those entitlements that can be transferred between buyers and sellers, inalienable entitlements are not considered.

2. *See infra* Part I.

3. *See The Cathedral*, *supra* note 1, at 1105.

4. *See, e.g.*, Parker v. Swett, 205 P. 1065 (Cal. 1922).

5. *See* Susan F. French, *Servitudes Reform and the New Restatement of Property: Creation Doctrines and Structural Simplification*, 73 CORNELL L. REV. 928, 928 (1988) (quoting EDWARD H. RABIN, FUNDAMENTALS OF MODERN REAL PROPERTY LAW 489 (1974)) [hereinafter French, *Servitudes*]. *See also* Susan F. French, *Toward a Modern Law of Servitudes: Reweaving the Ancient Strands*, 55 S. CAL. L. REV. 1261, 1261 (1982) (describing the law of servitudes as "com[ing] apart at the seams") [hereinafter French, *Reweaving*].

servitude law and went to great lengths reexamining and revising it.[6] During the revision, the drafters struck the use of pure property standards from certain servitude doctrines and instead applied liability rules.[7]

Not all aspects of servitude law, however, were placed under a microscope. One doctrine that has yet to be reviewed is the rule that easements may not be extinguished by nonuse.[8] Though nonuse may be one relevant factor in determining whether an easement can be terminated, nonuse alone has been held to be insufficient to abolish an easement for centuries by both the English and American common law.[9] Under the framework of Calabresi and Melamed, this longstanding rule against termination by nonuse constitutes a pure property rule.[10]

In contrast to the common law rule is the rule regarding nonused easements in civil law jurisdictions.[11] Since early Roman law, civilian systems have automatically terminated easements that are not used for a statutorily set period of time.[12] The civil law rule represents what might be considered a quasi-property rule under the Calabresi and Melamed structure.[13]

Despite the longevity of both the civil law rule in favor of termination by nonuse and the common law rule against termination by nonuse, courts and scholars in both legal systems have offered little justification for why nonused easements should be governed by property or quasi-property standards.[14] The shortage of explanation is perplexing given the difference in each system's rule, but the lacking rationale becomes more troubling upon the realization that the rules regarding nonused easements in both

---

6. *See* RESTATEMENT (THIRD) OF PROP.: SERVITUDES ch. 1, intro. note (2000).

7. *See, e.g.*, John A. Lovett, *A Bend in the Road: Easement Relocation and Pliability in the New Restatement (Third) of Property: Servitudes*, 38 CONN. L. REV. 1, 5–6 (2005).

8. *See infra* Part II.A. Under the *Restatement (Third) of Property*, no servitude arrangement—be it an easement, profit, or covenant—may be terminated by nonuse alone. RESTATEMENT (THIRD) OF PROP.: SERVITUDES § 7.4 cmt. c (2000). This article, however, focuses solely on easements.

9. Parker v. Swett, 205 P. 1065, 1067 (Cal. 1922); Clark v. Redlich, 305 P.2d 239 (Cal. Ct. App. 2d 1957); White v. Crawford, 10 Mass. (1 Tyng) 183, 189 (Mass. 1813); Hofmeister v. Sparks, 660 N.W.2d 637, 641 (S.D. 2003); Mueller v. Hoblyn, 887 P.2d 500, 505 (Wyo. 1994); Ward v. Ward, (1852) 7 Exch. 838, 839 (U.K.). The rule against termination by nonuse has also been promulgated by all three Restatements. *See* RESTATEMENT (THIRD) OF PROP.: SERVITUDES § 7.4 cmt. c (2000); RESTATEMENT (FIRST) OF PROP.: SERVITUDES § 504 cmt. d (1944).

10. *See infra* Part II.A.

11. In referring to jurisdictions based on the civil law, this article only refers to jurisdictions based on French civil law; jurisdictions based on German civil law are not considered.

12. *See infra* Part III.A.

13. *Id.*

14. *See generally supra*, notes 7–9.

systems allow for the inefficient use of property.[15] Given that a tenet of American property law is that property should be used efficiently,[16] it is puzzling how a doctrine that potentially encourages waste has stood unscathed for so long.

This article strives to examine nonused easements under the Calabresi and Melamed framework and evaluate what type of rule should be applied to nonused easements to promote the continuous efficient use of real property. Before scrutinizing nonused easements, Part I of this article provides a brief overview of *The Cathedral*'s framework. Then, the article uses that framework to closely examine the common law and civil law doctrines concerning nonused easements. Part II analyzes the common law rule against termination by nonuse, identifies why the rule is a property standard, and demonstrates how the application of the property standard creates inefficiencies. Part III follows the same approach for the civil law rule in favor of termination by nonuse: it surveys the civil law rule, establishes the quasi-property qualities of the rule, and identifies the strengths and weaknesses of applying a quasi-property rule to nonused easements.

Reflecting upon the shortcomings of applying a property or quasi-property rule, the article then examines nonused easements from a different perspective, a liability lens. To look at nonused easements through a liability lens, Part IV of this article introduces a gloss to the Calabresi and Melamed framework recently developed by legal scholars Abraham Bell and Gideon Parchomovsky: pliability rules.[17] The Bell and Parchomovsky gloss questions the static application of the Calabresi and Melamed framework and argues that property and liability rules can be applied in a dynamic fashion as pliability rules.[18] Under the Bell and Parchomovsky theory, when a triggering event occurs, the law can change what type of rule is applied.[19] Part V proposes applying pliability rules to easements once they fall into a state of nonuse. The pliability rule recommended by this article is a system of private eminent domain based on nonuse. Under this theory, the holder of the servient estate may terminate a nonused

---

15. *See infra* Part II.B; Part III.B.

16. Efficient use of property is the underlying idea behind a host of American property law institutions, including the rule against unreasonable restraints on alienation, the doctrine of adverse possession, and the doctrine of waste. *See* McInerny v. Slights, 1988 WL 34528, at *6 (Del. Ch. Apr. 13, 1998) (rule against unreasonable restraints on alienation); Hillsmere Shores Improvement Ass'n v. Singleton, 959 A.2d 130, 144 (Md. Ct. Spec. App. 2008) (doctrine of adverse possession); Devins v. Borough of Bogota, 592 A.2d 199, 202 (N.J. 1991) (citing RICHARD POSNER, ECONOMIC ANALYSIS OF LAW 70–71 (3d ed. 1986) (doctrine of adverse possession)); RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 74–75 (7th ed. 2007) (doctrine of waste).

17. *See* Abraham Bell & Gideon Parchomovsky, *Pliability Rules*, 101 MICH. L. REV. 1, 25–26 (2002).

18. *Id.* at 5.

19. *Id.* at 31–32.

easement if certain conditions are met, and, in turn, the holder of the nonused easement receives just compensation. The details of such a system, possible criticisms of its application, and responses for those critiques are contained in Part V. The article ends by questioning whether the concept of private eminent domain based on nonuse might be applied to other real property doctrines.

PART I. THE CALABRESI AND MELAMED FRAMEWORK

In *The Cathedral*, Calabresi and Melamed created a new framework for examining entitlements.[20] Their construct made two great contributions to legal scholarship. First, it provided a means of classifying rules to protect entitlements, and second, it gave guidance as to when each entitlement-protecting rule should be employed.[21]

In essence, Calabresi and Melamed crafted two lenses through which entitlement protections may be viewed: property standards and liability rules.[22] Property standards refer to entitlement protections in favor of the entitlement holder, while liability rules favor the nonholder of the entitlement.[23] Property standards protect the holder of the entitlement because they require the parties to enter into a voluntary transaction to alter the entitlement; under a property rule, the nonholder may only amend the arrangement if the holder of the entitlement agrees.[24] Liability rules, on the other hand, favor the nonholder of the entitlement because under a liability scheme, the nonholder can extinguish the entitlement by paying an objectively determined value.[25] Thus, under a liability rule, unilateral termination by the nonholder is allowed, provided the nonholder compensates the holder for that termination.

To illustrate how property rules and liability operate, Calabresi and Melamed used a hypothetical landowner and polluting factory based on the not-so-hypothetical case *Boomer v. Atlantic Cement Co.*[26] The scholars then described a dichotomy of legal remedies based on whether the landowner or the factory held the entitlement and what type of rule was applied.[27] When

---

20. *The Cathedral*, *supra* note 1, at 1092.

21. *See generally* Richard A. Epstein, *A Clear View of* The Cathedral: *The Dominance of Property Rules*, 106 YALE L.J. 2091 (1997) [hereinafter Epstein, *Clear View*]; Saul Levmore, *Unifying Remedies: Property Rules, Liability Rules, and Startling Rules*, 106 YALE L.J. 2149 (1997); Carol M. Rose, *The Shadow of* The Cathedral, 106 YALE L.J. 2175 (1997); Emily Sherwin, *Introduction: Property Rules as Remedies*, 106 YALE L.J. 2083 (1997).

22. *The Cathedral*, *supra* note 1, at 1105–12.

23. *Id.* at 1092.

24. *Id.*

25. *Id.*

26. Boomer v. Atlantic Cement Co., 257 N.E.2d 870 (N.Y. 1970).

27. *The Cathedral*, *supra* note 1, at 1115–16.

the entitlement of being free from pollution was held by the landowner, two possible rules resulted: Rule 1, a property rule, gave the landowner the right to unilaterally stop the factory from polluting; Rule 2, a liability rule, allowed the factory to continue polluting so long as the factory paid damages to the landowner.[28] When the entitlement of having the right to pollute was held by the factory, two different rules resulted: Rule 3, a property rule, allowed the factory to freely pollute without making any payment to the landowner; Rule 4, a liability rule, gave the landowner the ability to stop the pollution by paying the polluter just compensation.[29] With these four rules, Calabresi and Melamed created a complete framework of how entitlements could be protected.[30]

Understanding how entitlements could be protected allowed Calabresi and Melamed to determine when each set of rules should be employed.[31] Their scholarship suggests that efficiency is the predominant factor in determining when each rule should apply.[32] *The Cathedral* recommends the use of liability rules if the market valuation of the entitlement is inefficient or simply unavailable.[33] On the other hand, property standards should be utilized when transactions between the holder and nonholder are relatively cost-free and values can readily be assigned.[34] In such situations, parties are presumed to be able to organize a deal among themselves, so liability rules—which inherently require the involvement of a neutral third party arbitrator, such as a judge—are unnecessary.[35] In short, *The Cathedral* asserts that property standards should govern the entitlement when the party that can most efficiently bear the cost of entitlement is known with certainty because the market will encourage the person who can bear the entitlement's cost most efficiently to do so.[36] However, when the person who can most efficiently bear the costs is unknown, liability rules should be applied because a neutral third party will be necessary to insure that freeloaders and holdouts do not disrupt the market.[37]

In theory, *The Cathedral*'s rule is simplistically elegant;[38] in practice, it is frequently difficult to ascertain who can bear the costs most efficiently.[39]

---

28. *Id.*

29. *Id.* at 1116.

30. *Id.*

31. *See generally The Cathedral*, *supra* note 1, at 1090–93.

32. *Id.* at 1110. Calabresi and Melamed also suggested that society's distributive goals should guide when property standards or liability rules are used. *See id.* This article, however, focuses on the efficiency arguments made in *The Cathedral*.

33. *Id.* at 1110.

34. *Id.* at 1118.

35. *See generally id.*

36. *Id.* at 1119.

37. *Id.*

38. *See* Guido Calabresi, *Remarks: The Simple Virtues of the Cathedral*, 106 YALE L.J. 2201, 2202 (1997).

39. *The Cathedral*, *supra* note 1, at 1119.

Calabresi and Melamed highlighted this dilemma through their discussion of the hypothetical polluting factory and neighboring landowner.[40] Abstract guesses may be made as to whether the polluter or pollutee is in the best position to handle the costs of the pollution, but realistically, such a determination is based on a host of factors such as size of the parties, number of parties, amount of pollution, cost of alternatives, harm of pollution, etc. Even if every possible relevant factor was reasonably known, the status of the parties may change over time, thus shifting who is the best bearer of the burden.[41]

PART II. THE INEFFICIENCY OF LOOKING THROUGH A PROPERTY LENS: THE AMERICAN COMMON LAW RULE AGAINST TERMINATION BY NONUSE

Though the practical application of the Calabresi and Melamed framework may be difficult, it nonetheless remains a useful tool in understanding how society protects particular entitlements and whether the method applied is efficient. As such, scholars routinely rely on *The Cathedral*'s structure to evaluate legal doctrines.[42] One doctrine that scholars have yet to scrutinize is the common law rule against terminating easements by mere nonuse—a classic property rule. Though nonused easements appear to have some of the characteristics that, according to Calabresi and Melamed, make them suitable for being governed by a property regime, the use of property rules creates inefficiencies.[43]

A. The Common Law Rule Against Termination by Nonuse—a Classic Property Rule

An easement is a servitude that "creates a nonpossessory right to enter and use land in the possession of another."[44] Additionally, an easement establishes an obligation that the holder of the land over which the easement is granted may not interfere with the uses authorized by the

---

40. *Id.* at 1118–22.

41. That a change in the type of rule applied may occur is the crux of the Bell and Parchomovsky gloss. *See infra* Part IV.

42. *See, e.g.*, Ariel L. Bendor, *Prior Restraint, Incommensurability, and the Constitutionalism of Means*, 68 FORDHAM L. REV. 289, 311–18 (1999) (using the Calabresi and Melamed framework to examine free speech); Thomas F. Cotter, *Fair Use and Copyright Overenforcement*, 93 IOWA L. REV. 1271, 1292–99 (2008) (using the Calabresi and Melamed framework to examine copyright law); Lovett, *supra* note 7 (using the Calabresi and Melamed framework to examine easement relocation).

43. *See infra* Part II. A–B. *See generally The Cathedral*, *supra* note 1 (advocating that property rules should be employed when transaction costs are low).

44. RESTATEMENT (THIRD) OF PROP.: SERVITUDES § 1.2(1) (2000). Servitudes are a "legal device that creates a right or an obligation that runs with land or an interest in land." *Id.* § 1.1.

easement agreement.[45] In such a relationship, the burdened estate is referred to as the servient estate and the land that is benefitted by the arrangement is called the dominant estate.[46]

Like most property doctrines, the rules governing easements encourage efficiency in land transfers.[47] These rules achieve this goal by running with the land, meaning that future successors to the servient estate and dominant estates are automatically benefitted and burdened by the easement.[48] Automatic transfers of easements are considered efficient because they relieve successors from re-bargaining for the same, previously established easements, thereby removing repetitive transaction costs.

There are, however, instances in which one or both parties may prefer to end the easement rather than maintain the arrangement. To accommodate such situations, the American common law provides a variety of methods by which private parties[49] may terminate the existing easement arrangement: express agreement;[50] expiration of term;[51] release;[52] abandonment;[53] merger;[54] estoppel;[55] prescription;[56] and changed conditions.[57] Termination by nonuse is absent from this list.

---

45. *Id.* § 1.2(1).

46. EMORY WASHBURN, A TREATISE ON THE AMERICAN LAW OF EASEMENTS AND SERVITUDES 3 (Beard Books, 2d ed. 2000) (1867). Under American property law, easements may also be appurtenant or in gross. RESTATEMENT (THIRD) OF PROP.: SERVITUDES § 1.5 (2000). An appurtenant easement is one that is attached to the ownership and enjoyment of a particular tract of land, whereas an easement in gross is a personal right that is not attached to the benefit of a particular tract of land. *See* French, *Reweaving*, *supra* note 5, at 1267–68. For the purposes of this article, whether an easement is appurtenant or in gross is irrelevant. Thus, this article refers interchangeably to easement benefits for dominant estate (appurtenant) and easement benefits for a particular, personal beneficiary (in gross).

47. *See infra* Part II.B; Part III.B. In drafting the *Restatement (Third) of Property: Servitudes*, the drafters tried to design rules that allowed for the extinguishment of an easement when it became "obsolete, economically wasteful, or unduly burdensome." *See generally* French, *Reweaving*, *supra* note 5, at 1313. In other words, the drafters attempted to construct modes of termination that maintained efficient easements, but abolished inefficient easements.

48. Day v. Buckeye Water Conserv. & Drainage Dist., 237 P. 636, 640 (Ariz. 1925); RESTATEMENT (THIRD) OF PROP.: SERVITUDES § 1.1 (2000); WASHBURN, *supra* note 46, at 3; LEONARD A. JONES, A TREATISE ON THE LAW OF EASEMENTS § 1 (Baker, Voorhis & Co., 1898); French, *Servitudes*, *supra* note 5, at 928.

49. This article is specifically limited to the termination of easements by private parties. While some of the ideas expressed herein are modeled after methods by which the government may abolish property rights, this article only considers private parties acting to abolish certain property rights.

50. RESTATEMENT (THIRD) OF PROP.: SERVITUDES § 7.1 (2000).

51. *Id.* § 7.2.

52. *Id.* § 7.3.

53. *Id.* § 7.4.

54. *Id.* § 7.5.

55. *Id.* § 7.6.

Simply put, nonuse does not abolish an easement in the American common law.[58] This is true whether the easement lays unused for sixteen,[59] twenty-seven,[60] thirty-six,[61] forty,[62] or even 170 years.[63] Once an easement is conveyed to a dominant estate, the dominant estate holder and his successors in title will maintain an interest in that easement regardless of whether the easement is used.[64] The easement may never be used or may temporarily be used and then not used. Either way, the same rule prevails: easements are not terminable by nonuse.

While nonuse alone is not a recognized means of terminating easements, it is a factor in two accepted modes of extinguishment: abandonment and prescription.[65] To end an easement through either of these methods, the holder of the easement must not use his easement *and* some additional act must take place.[66] In the context of abandonment, the holder of the dominant estate performs the additional act;[67] in the context of prescription, the additional act is performed by the holder of the servient estate.[68]

To abandon an easement, the beneficiary of the easement must intentionally relinquish the rights created by the easement.[69] There is no formal instrument that must be executed for the dominant estate holder to abandon his easement.[70] Because of the lack of formality required, if the

---

56. *Id.* § 7.7.

57. *Id.* § 7.10.

58. *Id.* § 7.4 cmt. c. In discussing the rule against termination by nonuse, this article inherently limits the notion of easements to affirmative easements, which are easements in which the dominant estate holder is given the right to perform some act on the servient estate holder's land. How a negative or spurious easement might be terminated by nonuse is conceptually unclear in the common law, though under the civil law framework, the dominant estate may lose what the common law would consider a negative or spurious easement due to nonuse.

59. *See* Carnemella v. Sadowy, 538 N.Y.S.2d 96 (N.Y. App. Div. 1989).

60. *See* Mueller v. Hoblyn, 887 P.2d 500 (Wyo. 1994).

61. *See* White v. Crawford, 10 Mass. (1 Tyng) 183 (Mass. 1813).

62. *See* Arnold v. Stevens, 41 Mass. (1 Pick.) 106 (Mass. 1839).

63. *See* Pencader Assoc. v. Glasgow Tr., 446 A.2d 1097, 1101 (Del. 1982).

64. *See* McCulloch v. Roberts, 276 So. 2d 425 (Ala. 1973); Castle Assoc. v. Schwartz, 407 N.Y.S.2d 717 (N.Y. App. Div. 1978).

65. *See* RESTATEMENT (THIRD) OF PROP.: SERVITUDES § 7.4 cmt. c (2000); *id.* § 7.7 cmt. b.

66. *See id.* § 7.4 cmt. c; *Id.* § 7.7 cmt. c.

67. *See id.* § 7.4 cmt. c.

68. *See id.* § 7.7 cmt. c.

69. Bank of Fayetteville v. Matilda's, Inc., 803 S.W.2d 549, 550 (Ark. 1991) (citations omitted); Miller v. St. Louis, Sw. Ry. Co., 718 P.2d 610, 613 (Kan. 1986) (citations omitted); Chase v. Eastman, 563 A.2d 1099, 1102 (Me. 1989).

70. The lack of formal instrument is what separates abandonment from release. *Cf.* RESTATEMENT (THIRD) OF PROP.: SERVITUDES § 7.3 cmt. a ("A release is the method ordinarily used to effectuate a formal extinguishment of rights of a servitude beneficiary.").

holder of the burdened estate wishes to demonstrate that the easement is abandoned, the beneficiary's intent to abandon must be proved.[71] Evidence of intent generally requires reliance on circumstantial evidence[72] that is "decisive and conclusive."[73] To meet such a burden, nonuse frequently serves as one piece of evidence indicating the dominant estate holder's intent to abandon.[74] The length of the nonuse may affect the amount of evidence required to prove abandonment—the longer the nonuse, the less evidence needed.[75] Though an elongated period of nonuse may necessitate less evidence, providing *some* additional evidence is crucial, as was demonstrated in *Castle Associates v. Schwartz*.[76] In *Castle*, the owner of the servient estate granted an easement to the owner of the dominant estate in 1903.[77] The easement was never used.[78] In 1976, the existence of the "forgotten easement" was revealed through a title search performed by successors-in-title to the original dominant estate holder.[79] The New York court held that despite the seventy-three years of nonuse, the burdened landowner failed to prove abandonment because he did not offer additional evidence signaling that the benefitted landowner intended to abandon the easement.[80]

---

71. W. Union Tel. Co. v. Louisville & N. R. Co., 81 So. 44, 50–51 (Ala. 1918); Am. Brass Co. v. Serra, 132 A. 565, 568 (Conn. 1926); Knotts v. Summit Park Co., 126 A. 280, 282 (Md. 1924).

72. Direct expressions of an intent to abandon are rarely available. "[A] servitude beneficiary who deliberately sets about divesting him or herself of a servitude interest normally uses a release." RESTATEMENT (THIRD) OF PROP.: SERVITUDES § 7.4 cmt. a.

73. Richardson v. Tumbridge, 149 A. 241, 242–43 (Conn. 1930); Tietjen v. Meldrim, 151 S.E. 349, 359–60 (Ga. 1930); Hayford v. Spokesfield, 100 Mass. 491, 494 (Mass. 1868). *See* Phillips v. Gregg, 628 A.2d 151, 153 (Me. 1993); *Miller*, 718 P.2d at 613 (citations omitted).

74. McCulloch v. Roberts, 276 So. 2d 425, 428 (Ala. 1973); Mariano v. Guarino, No.0101555, 1993 WL 117741, at *3 (Conn. Super. Ct. 1993) (quoting Byard v. Hoelscher, 16 151 A. 351 (1930)); Millson v. Laughlin, 142 A.2d 810, 816–17 (Md. 1958) (citing 3 TIFFANY, REAL PROPERTY § 825 (3d ed. 1939)). A Georgia statute provides that "[a]n easement may be lost by abandonment or forfeited by nonuse if the abandonment or nonuse continues for a term sufficient to raise the presumption of release or abandonment." GA. CODE ANN. § 44-9-6 (2008). Georgia courts have interpreted this statute to mean that after twenty years of nonuse there is a rebuttable presumption of abandonment. *See* Duffy St. S.R.O., Inc. v. Mobley, 471 S.E.2d 507, 508 (Ga. 1996); Gilbert v. Reynolds, 212 S.E.2d 332, 335 (Ga. 1975); Kelsoe v. Oglethorpe, 48 S.E. 366, 368 (Ga. 1904). Some early jurisprudence from other states exists that echoes the Georgia statute. *E.g.*, City of Peoria v. Johnston, 56 Ill. 45, 49 (Ill. 1870); Corning v. Gould, 16 Wend. 530, 535–36 (N.Y. Sup. Ct. 1837).

75. RESTATEMENT (THIRD) OF PROP.: SERVITUDES § 7.4 cmt. c.

76. *See* Castle Assoc. v. Schwartz, 407 N.Y.S.2d 717 (N.Y. App. Div. 1978).

77. *Id.* at 719.

78. *Id.* at 720.

79. *Id.*

80. *Id.* at 721.

Whereas termination by abandonment concentrates on the conduct of the holder of the dominant estate, termination by prescription focuses on the actions of the holder of the servient estate.[81] To terminate an easement by prescription, the burdened party must demonstrate that he adversely used the easement,[82] and that his use was open, notorious, and continuous without interruption for the statutorily mandated period of time.[83] Nonuse impacts prescription because by the beneficiary's nonuse, the servient estate holder is able to possess the easement continuously without interruption.[84] If the dominant estate holder uses his easement amid the adverse possession, the running of prescription ceases because the adverse use of the easement by the burdened party is no longer continuous and uninterrupted.[85] As such, nonuse by the beneficiary is necessary to prove termination by prescription, but that nonuse must be accompanied by the adverse use of the easement by the holder of the servient estate.[86]

While nonuse may be relevant to determining whether an easement should be extinguished, nonuse by itself does not terminate an easement at common law.[87] This doctrine represents a classic property rule within the

---

81. Matoush v. Lovingood, 177 P.3d 1262, 1270 (Colo. 2008); *see* RESTATEMENT (THIRD) OF PROP.: SERVITUDES § 7.4 cmt. b (2000).

82. Not just any use constitutes an adverse use. In order for use of the easement by the owner of the servient estate to be adverse, the use "must be incompatible or irreconcilable with the easement holder's right to use the easement." *Matoush*, 177 P.3d at 1265. *Accord* Peasel v. Dunakey, 279 S.W.3d 543, 546 (Mo. Ct. App. 2009); Rowe v. Lavanway, 904 A.2d 78, 84 (Vt. 2006).

83. *Matoush*, 177 P.3d at 1265; Creech v. Noyes, 87 S.W.3d 880, 885 (Mo. Ct. App. 2002) (citations omitted). *See also* Mueller v. Hoblyn, 887 P.2d 500 (Wyo. 1994) (distinguishing between prescription extinguishing ownership and prescription extinguishing easements).

84. *See generally* RESTATEMENT (THIRD) OF PROP.: SERVITUDES § 7.7 (2000).

85. *See generally id.*

86. *See generally id.*

87. *Supra* notes 58–63. Fifteen jurisdictions have either legislatively or judicially adopted an exception to this rule: servitudes that are acquired by prescription can be extinguished by nonuse for the length of time necessary to acquire the easement by prescription. The jurisdictions that have adopted this exception include: Arizona, Arkansas, California, Guam, Maine, Michigan, New York, North Carolina, North Dakota, Oklahoma, Pennsylvania, South Carolina, South Dakota, Tennessee, and Wisconsin. The exceptions in California, Guam, North Dakota, and Oklahoma are statutorily based exceptions, whereas the other jurisdictions have judicially based exceptions. *See* CAL. CIV. CODE § 811 (West 2009); N.D. CENT. CODE § 47-05-12 (West 2009); 4 OKLA. STAT. tit. 60, § 59 (2009); 21 GUAM CODE ANN. § 7111 (2009); Johnston v. Verboon, 598 S.W.2d 752 (Ark. 1980); Furrh v. Rothschild, 575 P.2d 1277 (Ariz. Ct. App. 1978); Adams v. Hodgkins, 84 A. 530 (Me. 1912); McDonald v. Sargent, 13 N.W.2d 843 (Mich. 1944); Miller v. Garlock, 8 Barb. 153 (N.Y. App. Div. 1850); Woodbury v. Allan, 64 A. 590 (Pa. 1906); Nitzell v. Paschall, 3 Rawle 76 (Pa. 1831); Monaghan v. Memphis Fair & Exposition Co., 31 S.W. 497 (Tenn. 1895); Parkins v. Dunham, 34 S.C.L. (3 Strob.) 224 (S.C. App. L. 1848); Shippy v. Hollopeter, 304 N.W.2d 118 (S.D. 1981); Johnson v. Boorman, 22 N.W. 514 (Wis. 1885).

Calabresi and Melamed framework because the holder of the easement is protected from the nonconsensual extinguishment of the easement by the nonholder of the easement. In other words, the burdened party cannot unilaterally remove the easement from his estate; he must obtain the beneficiary's consent to rid his estate of the easement.[88]

B. The Inefficiency in Applying a Property Rule

According to Calabresi and Melamed, property rules should be employed when transaction costs are low.[89] Given that the parties involved in a nonused easement are easily identifiable and usually small in number, it appears at first glance that transaction costs for altering nonused easements are relatively low, thus making property rules the ideal type of rule to govern the situation.[90] However, by closely observing how the rule against termination by nonuse affects nonused easements, it becomes clear that applying a property standard to the doctrine creates inefficiency.[91] This inefficiency is created because of the parties' inability to predict the future, the societal loss created, the encouragement of speculative purchasing, and the possibility of holdups by the dominant estate.

1. Inability to Predict the Future

When the holders of a servient and dominant estate enter into an easement arrangement, the price of the easement should reflect the amount that both parties value the easement. For example, suppose that Sara owns Blackacre in fee simple and grants an easement for $100 to David. David's easement consists of a right to drive across Blackacre. At the time in which Sara, the servient estate holder, and David, the dominant estate holder, enter into this agreement, it can be presumed that they both value the easement at $100. How each determines his personal value of the easement differs between the parties.

David's valuation of the easement is determined by his predicted use of the easement. Suppose that instead of driving across Blackacre, David could take an alternate route and still reach the same destination point. David will then value each individual use of the easement across Blackacre

---

While this exception theoretically exists, it is rarely successfully employed.

88. The only means by which the servient estate holder may abolish the easement without the consent of the dominant estate holder is through prescription. However, even prescription arguably requires the consent of the dominant estate holder, in that he can at any point in time stop the running of the prescriptive period by simply using the easement. *See* Castle Assoc. v. Schwartz, 407 N.Y.S.2d 717, 721; RESTATEMENT (THIRD) OF PROP.: SERVITUDES § 7.7 (2000).

89. *The Cathedral*, *supra* note 1, at 1118.

90. *See generally* RESTATEMENT (THIRD) OF PROP.: SERVITUDES § 7.4 cmt. a (2000).

91. *See generally* RESTATEMENT (THIRD) OF PROP.: SERVITUDES § 7.1 (2000).

at the cost he would have to pay to take the alternate route. If taking an alternate route costs David $1, then every use David makes of the easement is worth $1 to him. David's value of the easement should naturally factor in how many times he expects to use the easement and the cost for each use. Therefore, when David agrees to pay $100 for the easement, it can be presumed that he expects he will use the easement at least 100 times.

Sara's value of the easement is similarly based on her next-best option, though Sara's opportunity cost refers to what else she could do with Blackacre. If the next best thing Sara could do with Blackacre is farm wheat, then Sara must predict the amount of wheat she will be unable to farm due to David's easement. When Sara charges David $100, $100 should equal what Sara expects she could have gained had she used Blackacre to farm wheat. Thus, in a perfect market, Sara and David will determine their individual values for the easement and will only enter into an easement arrangement if their individual values for the easement align. If David thinks he will use the easement only 100 times at $1 per use, but Sara believes she can grow $1,000,000 worth of wheat on the land where David desires his easement, there will be no deal because Sara will be unwilling to sell the easement for $100 and David will be unwilling to pay the $1,000,000 Sara will demand.

The problem with subjecting easements to a property rule is that David and Sara are temporally constrained in their ability to predict the amount they value the easement.[92] David may be able to predict the amount that he will use the easement during the subsequent month or even year, but any prediction David makes about the amount of his use of the easement ten years down the road is nothing more than speculation. [93] Similarly, the ability of Sara to predict her opportunity cost for granting the easement is time limited. Moreover, neither party will know the value of the easement arrangement to their successors.[94]

---

92. Professor Richard A. Epstein argues that though contracting parties may not be able to accurately predict the future, they are "aware of the difficulties of dealing with future uncertainty," and contract accordingly. Richard A. Epstein, *Covenants and Constitutions*, 73 CORNELL L. REV. 906, 924 (1988) [hereinafter Epstein, *Covenants*]. Professor Epstein may be correct that parties to servitude arrangements take into account the uncertain future when entering into the original arrangement. The problem with the strict property rule is that it never allows courts to step in to remedy issues that arise. But, even Professor Epstein admits that "the parties' inability to draft with perfect foresight and completeness necessarily means that courts will have to engage in some 'interstitial legislation' in construing the terms of the basic agreement." *Id.* at 923 (footnote omitted). However, the rule at issue here, the rule against termination by nonuse, does not allow for such "interstitial legislation." *Id.* at 923.

93. For a discussion of how speculative purchasing creates inefficiencies, see *infra* Part II.B.3.

94. Professor Gerald Korngold writes, "The current, most efficient use of land is hard enough to determine; trying to predict how land should best be used by future generations requires a good measure of hubris." Gerald Korngold, *Resolving the Intergenerational Conflicts of Real Property Law: Preserving Free Markets and Personal Autonomy for*

Parties routinely enter into property arrangements requiring them to (unrealistically) predict the future. In the easement context, however, imposing a property rule like the rule against termination by nonuse that requires Sara and David to speculate about the value of their future respective uses is particularly problematic because, as conditions change, Sara, the servient estate holder, is prevented from altering the easement arrangement, but David is not. If the easement somehow becomes a burden on David, then he may execute a release[95] or abandon it.[96] Or, if David and his successors simply have no desire to use the easement, they can easily cease their use. All of these options come at no cost to David. While David has an easy, cost-free method of escaping the easement agreement, Sara and her successors remain bound by the original agreement, regardless of whether David and David's successors stop using the easement entirely, thus indicating that they no longer value the easement. So long as the dominant estate holder does not take affirmative steps to release the servient estate holder or does not demonstrate an intent to abandon the easement, the servient estate remains burdened with the easement and the agreement established by the original valuations.[97] As such, Sara must continue not interfering with David's authorized use, despite the fact that he is never using the easement. In the hypothetical, this means that despite David's nonuse, Sara must continue to not grow wheat.

Theoretically, the concern that parties' valuation of an easement will change over time is addressed by the changed-conditions doctrine.[98] The rationale underlying the changed-conditions rule is to prevent obsolete servitudes from interfering with desirable uses of land.[99] However, because courts have looked at easements through a property lens, the changed-conditions doctrine has been applied with great hesitation[100] and some courts will simply not apply a changed-conditions doctrine to easements.[101] The test developed by the *Restatement (Third)* provides fairly restrictive

---

Future Generations, 56 AM. U. L. REV. 1525, 1553 (2007).

95. RESTATEMENT (THIRD) OF PROP.: SERVITUDES § 7.3 (2000).

96. *Id.* § 7.4.

97. *See id.*

98. *See id.* § 7.10.

99. *Id.* § 7.10 cmt. a.

100. *Id.* § 7.10 cmt. a; *see* Chevy Chase Vill. v. Jaggers, 275 A.2d 167, 171 (Md. Ct. App. 1971) (The change in condition must be "so radical as to render perpetuation of the restriction of no substantial benefit to the dominant estate, and to defeat the object or purpose of the restriction."). The very idea of a changed-conditions doctrine has been criticized by scholars because it allows for the interference of property rights. *E.g.*, Robert Ellickson, *Alternates to Zoning: Covenants, Nuisance Rules, and Fines as Land Use Control*, 40 U. CHI. L. REV. 681, 716–17 (1973); Richard A. Epstein, *Notice and Freedom of Contract in the Law of Servitudes*, 55 S. CAL. L. REV. 1353, 1364–68 (1982).

101. *See* Cortese v. United States, 782 F.2d 845, 851 (9th Cir. 1986) (implying the changed conditions doctrine only applies to covenants and not to easements); Waldorp v. Brevard, 62 S.E.2d 512, 515 (N.C. 1950).

criteria for when conditions are changed such that an easement may be terminated or modified:

> (1) When a change has taken place since the creation of a servitude that makes it *impossible* as a practical matter to accomplish the purposes for which the servitude was created, a court may modify the servitude to permit the purpose to be accomplished. If modification is not practicable, or would not be effective, a court may terminate the servitude.
> (2) If the purpose of a servitude can be accomplished, but because of changed conditions the servient estate is no longer suitable for uses permitted by the servitude, a court may *modify* the servitude to permit other uses under conditions designed to preserve the benefits of the original servitude.[102]

Under the *Restatement (Third)*, the easement may be terminated only when the easement is impossible and unable to be modified.[103] In the case of Sara and David, David's mere nonuse of the easement does not render the easement "impossible."[104] Thus, there is no reason why under the changed-conditions doctrine David's nonuse should automatically free Sara's estate of the easement.

Moreover, and more troubling, the changed-conditions doctrine does not come to Sara's aid when David stops using the easement *and* her opportunity costs change. Suppose instead of growing wheat for $100, now Sara can develop Blackacre into a mega shopping complex that would be worth $1,000,000. Even though David is still not using the easement and Sara would now place a much higher value on the easement, the changed-conditions doctrine does not operate to terminate the easement. The changed-conditions doctrine might help to modify the easement arrangement, but only if the benefits of the original easement can be preserved.[105] Thus, the narrowly drawn changed-conditions doctrine provides little to no benefit to the servient estate holder when the originally predicted use of the easement, and in turn, the value of the easement, changes.[106] As the use and value change, so too does the efficiency of applying a property standard.

---

102. RESTATEMENT (THIRD) OF PROP.: SERVITUDES § 7.10 (2000) (emphasis added).

103. *See id.*

104. *See id.*

105. *See* RESTATEMENT (FIRST) OF PROP. § 564 (1944).

106. While this article describes the changed conditions doctrine as "narrowly drawn," it must be noted that the current changed conditions doctrine is broader in scope than it used to be. *Cf.* RESTATEMENT (FIRST) OF PROP. § 564 (1944). In discussing the need to allow courts to at least minimally change contracts, Professor Stewart Sterk points out, "no one believes that contracting parties are blessed with perfect foresight." Stewart E. Sterk, *Foresight and the Law of Servitudes*, 73 CORNELL L. REV. 956, 957 (1988). To remedy the inadequate foresight of the parties, contract law allows numerous "escape valves" by which

2. Creation of Societal Loss

In some respects, the fact that the law does not provide a remedy for the servient estate holder in the case of a nonused easement is not troubling. While parties like Sara and David may not be able to accurately predict how much they will value the easement in the future, they still voluntarily entered into the easement arrangement. If Sara was truly concerned that she would be unable to foresee the possible future utility of Blackacre, she never should have entered into the agreement. When Sara conveyed the easement to David, she took the risk that over time, $100 would be less than her opportunity costs for Blackacre. Viewed in this light, perhaps the law should not provide a safety net for Sara simply because years down the road she (or her successors) realizes that she entered into a bad deal.[107]

The problem with this rationale is that more individuals than just Sara and her successors lose from her bad investment: society also loses from a nonused easement. If David stops using the easement, Sara still may not engage in activities that would prevent David from using the easement. In as much as farming wheat would prevent David from driving across Blackacre, Sara is precluded from doing it. Society loses from this preclusion because if Sara could farm wheat on Blackacre, she would purchase items like seed, fertilizer, plows, and tractors. She might employ field hands. At the end of the wheat-farming season, she would likely sell the wheat. All of these activities have a positive economic impact on society. Society's loss becomes more egregious as Sara's opportunity costs rise. If Blackacre can be converted from a tract of land suited for wheat growing to a tract of land suited for a mega shopping complex, then society, albeit indirectly, has lost out on that opportunity cost. So long as the easement runs across Blackacre, Sara will be unable to engage in such activities, thereby denying society the potential benefits.

Obviously society suffers the same loss when David uses the easement, for Sara is prevented from farming wheat then, too. However, when David uses the easement, the loss to society is less bothersome because it is balanced by David's gains. Every time David uses the easement, he receives a benefit from the use. Though society may be foregoing Sara's opportunity costs, David's gains from his use of the easement offset at least some of society's loss. Admittedly, David's gains may not be equivalent to what society stands to gain by utilizing Blackacre, but at least if David gains something, society's loss is more palatable. However, when David fails to use the easement, no one is gaining from the existence of the easement. Instead, everyone is only losing.

---

courts may remedy contracts. *Id.* at 961. The rule against termination by nonuse, however, is lacking in the escape valves highlighted by Professor Sterk.

107. Professor Epstein raises this same argument in his comparison of covenants and constitutions. *See generally* Epstein, *Covenants*, *supra* note 92.

To at least some extent, the rule that easements may be terminated by prescription mitigates the societal loss from David's nonuse. When David stops using the easement, Sara may adversely use the easement, thus commencing the running of prescription.[108] In other words, once David stops using the easement, Sara can farm her wheat (or develop her mega shopping complex). If David continues to not use the easement and Sara continues in her adverse, hostile, open, and notorious wheat growing, then at the end of the prescriptive period, David's easement will be extinguished.[109] In this sense, prescription acts as a self-help remedy for Sara because Sara can utilize David's nonuse of the easement to terminate David's easement without his express consent.

Though prescription lessens the potential loss to society due to David's nonuse, it is a far from perfect solution. David can interrupt and stop the running of the prescriptive period at any moment by simply using this easement. This makes attempting to terminate an easement by prescription quite risky for Sara because at any point, David may use the easement, thus quashing any investment Sara made. When presented with this type of situation, the chances that Sara will purposefully engage in an adverse use so as to terminate the easement are slim, particularly if the desired adverse use involves a large investment, such as building a mega shopping mall complex.[110]

3. Encouraging Speculation and Division in Land

Because the rule against termination by nonuse is a classic property rule, dominant estate holders like David are allowed—and arguably encouraged—to engage in speculative purchasing.[111] Suppose instead of intending to drive across Sara's land once a day, in reality David wishes to purchase the easement because he believes that one day, Blackacre will transform into an Eden-like garden that will allow David to achieve eternal bliss when he drives across it. Until that day arrives, however, David has no intention of utilizing the easement. In this case, David will buy the easement now and simply not use it until the day of bliss arrives.

The problem with allowing David to speculate about the future use of the easement is that because he has no ability to accurately predict the future,[112] his speculative purchase of the easement on Blackacre is a gamble at best. While the law generally allows individuals to gamble when entering

---

108. *Supra* Part II.A.

109. *See* RESTATEMENT (THIRD) OF PROP.: SERVITUDES § 7.4 cmt. b.

110. For an account of how bad-faith adverse possession may actually produce efficiency, see Lee Anne Fennell, *Efficient Trespass: The Case for "Bad Faith" Adverse Possession*, 100 NW. U. L. REV. 1037, 1095 (2006).

111. In contrast to the arguments raised in this article, Judge Posner has harshly critiqued comments regarding speculation as a negative. POSNER, *supra* note 16, at 49.

112. *See* Part II.B.1.

contracts, contract law also places restrictions on when the gambling party must ante up. For example, the law prohibits indefinite option contracts.[113] Instead, option contracts must be exercised in the time period provided by the parties, and if no time limitation is stated, within a reasonable time.[114] Courts have strictly enforced time limitations for option contracts because "any relaxation of terms would substantively extend the option contract to subject one party to greater obligations than he bargained for."[115] With easements, however, there is no restriction on when David must act on his speculation. Thus, an easement acts like a perpetual option.[116]

The reason that perpetual options are not allowed—because it would "give all advantage to the one in whose favor the option was granted"[117]—applies *a fortiori* to easements. Not only does a nonused easement force the servient estate holder to protect the dominant estate holder's right to use the easement more than is required, but it also requires the burdened estate to lie in a perpetual state of divided possession.[118] Though easements create a nonpossessory right in property,[119] given that the servient estate holder must refrain from interfering with the beneficiary's use—even when the beneficiary is not using the easement—the burdened estate is essentially divided into two separate possessory interests.[120] Sara may only possess portions of Blackacre to the extent that the possession does not interfere with David's use, and David may possess portions of Blackacre in

113. *See* Iglehart v. Philips, 383 So. 2d 610, 615 (Fla. 1980); Brine v. Fertitta, 537 So. 2d 113, 114 (Fla. Dist. Ct. App. 1988); H. Evans Scobee, *The Requirement of a Definite Time Period in Option Contracts*, 34 LA. L. REV. 668, 674–75 (1974); RESTATEMENT (SECOND) OF CONTRACTS § 87 (1981).

114. *See* Ex parte Keelboat Concepts, Inc. v. C.O.W., Inc., 938 So. 2d 922, 925 n.3 (Ala. 2005) (citations omitted); Hughes v. Holliday, 99 S.E. 301, 302–03 (Ga. 1919); Hermes v. William F. Meyer Co., 382 N.E.2d 841, 844–45 (Ill. App. Ct. 1978).

115. *See* RESTATEMENT (SECOND) OF CONTRACTS § 25 cmt. c (1981).

116. It is plausible that the "perpetual option" property will never be a problem. If David knows he will not use the easement until Blackacre transforms into Eden, then he may be inclined to inform Sara of his delayed use and allow Sara to farm wheat on Blackacre until the day of transformation arrives. Because Sara will be able to reap the benefits of Blackacre while David waits, David may hypothesize that Sara will demand less for the easement, thus lowering the contract price to David. However, Sara may alternatively realize that any transformation of Blackacre is a number of years away, so she may decide to wait to contract for the easement closer to the actual date because her opportunity costs for Blackacre may change. If David is concerned that the latter will occur, he will be incentivized to not tell Sara why he is purchasing the easement or about his anticipated delayed use.

117. Clark v. Dixon, 254 So. 2d 482, 483 (La. Ct. App. 1971).

118. Professors Ben W. F. Depoorter & Francesco Parisi view servitudes as "a partitioning of property rights." Ben W. F. Depoorter and Francesco Parisi, *Fragmentation of Property Rights: A Functional Interpretation of the Law of Servitudes*, 3 GLOBAL JURIST FRONTIERS art. 2, 19 (2003), *available at* http://www.bepress.com/gj/frontiers/vol3/iss1/art2/.

119. RESTATEMENT (THIRD) OF PROP.: SERVITUDES § 1.2 (2000).

120. *See id.*

accordance with his contracted-for use.[121] The bundle of possessory rights attached to Blackacre is divided indefinitely between Sara and David.[122]

Given the well-known maxim that possession is nine-tenths of the law,[123] the correlating maxim *nemo invitus ad communionem compellitur*[124] should apply equally to possessory divisions in land. Generally, the law disfavors perpetual division in the ownership of land to such a degree that partition by cotenants is considered a matter of right.[125] Agreements that perpetually limit the ability to partition real property are void because they prohibit the alienation of property.[126] Unpartitioned land interests impinge upon the sale of real estate because purchasing a portion of an interest in land is frequently less desirable than purchasing an entire interest in land.[127] Nonused easements produce a similar effect on the land they burden; estates with nonused easements are undesirable to buyers because the buyer will have to maintain protection of the dominant estate holder's right to use the easement without reaping any benefit from the easement.[128]

4. Giving Rise to Holdups

In addition to the difficulties caused by speculative purchasing, a further problem arises from the imposition of a property rule to nonused easements: the possibility of holdups. If Sara desires to rid Blackacre of the easement, she has only one option—she must strike a deal with David. Because David is the only person who can contract with Sara to rid Blackacre of the easement, David is able to hold Sara up for a price above his opportunity cost.[129] Suppose when David and Sara enter the easement

---

121. *See* Depoorter & Parisi, *supra* note 118, at 19–23.

122. *See generally id.*

123. *See generally* FREDERICK POLLOCK & ROBERT SAMUEL WRIGHT, AN ESSAY ON POSSESSION IN THE COMMON LAW 5 (Adamant Medi Corp. 2000) (1888).

124. No one can be forced to have common property with another. This concept has been applied in numerous cases. *See, e.g.*, Delfino v. Vealencis, 436 A.2d 27, 30 (Conn. 1980); Hall v. Hamilton, 667 P.2d 350, 354–55 (Kan. 1983); Ark Land Co. v. Harper, 599 S.E.2d 754, 758–59 (W. Va. 2004).

125. Watson v. Durr, 379 So. 2d 1243, 1244 (Ala. 1980); Hamilton v. Hamilton, 597 A.2d 856, 859 (Del. Fam. Ct. 1990); Chew v. Sheldon, 108 N.E. 552, 552 (N.Y. 1915).

126. Albin v. Albin, 208 N.Y.S.2d 252, 256 (N.Y. App. Div. 1960).

127. *See generally* Francesco Parisi, *Freedom of Contract and the Laws of Entropy*, 10 SUP. CT. ECON. REV. 65 (2002) (discussing the dysfunctional fragmentation of real property).

128. *See generally* RESTATEMENT (THIRD) OF PROP.: SERVITUDES § 1.2 (2000) (explaining the rights associated with easements).

129. Professors Thomas Miceli and C.F. Sirmans point out that:

> [A] true holdout problem requires assembly. . . . [N]egotiations between a buyer and seller for a single parcel . . . do not constitute a holdout problem . . . because the seller's unwillingness to sell does not affect any other transactions; it simply reflects his efforts to obtain the highest possible price.

arrangement, they both have the opportunity cost of $100. If Sara's opportunity cost increases to $300 while David's opportunity cost remains the same, then in a free market Sara should be able to pay David $100 to rid Blackacre of the easement. By paying David $100, Sara is made better off and David is left no worse off. Thus, Sara's backing out of the original easement agreement, or "breaching" the agreement, is efficient.[130]

David, however, will not charge Sara $100, but instead will hold out for some price between $100 and $200.[131] Because David is the only person with whom Sara can realistically contract for a release of the easement, Sara will pay anything below her increased opportunity cost.[132] While the holdup by David may not cause inefficiency per se,[133] it does alter the price that the easement would have sold for if a free market existed.

The holdup situation is augmented when David stops using his easement. Suppose that David stops using his easement because he no longer needs to drive across Blackacre to reach his destination. The value of the easement to David is now much less than it was when he needed to drive across Blackacre. If David now has no value for the easement, so a value of $0, then in the free market, he would give the easement away. But

---

Thomas J. Miceli & C.F. Sirmans, *The Holdout Problem, Urban Sprawl, and Eminent Domain* 5 (Nov. 2006), http://ssrn.com/abstract=952511 (follow "One-Click Download" hyperlink to download SSRN-id952511.pdf). Based on the law and economic scholars' description, it may appear that in the case of David and Sara, no holdup can occur because David is the only seller in the market. However, Sara does have another option; she may purchase a different tract of land without an easement to pursue whatever activity she wishes to engage in but cannot because of David's easement. As such, there is an assembly of potential sellers—David and all sellers of land. However, while Sara does technically have other options, David has a distinct selling advantage and can hold up Sara for an above-market price.

130. *See generally* Melvin A. Eisenberg, *Actual and Virtual Specific Performance, the Theory of Efficient Breach, and the Indifference Principle in Contract Law*, 93 CAL. L. REV. 975, 997–1016 (2005) (examining the theory of efficient breach); Ronald J. Scalise Jr., *Why No "Efficient Breach" in the Civil Law?: A Comparative Assessment of the Doctrine of Efficient Breach of Contract*, 55 AM. J. COMP. L. 721 (2007) (evaluating the American theory of efficient breach of contract).

131. Professor Richard Epstein posits a similar hypothetical in his argument that liability rules should be utilized when a holdout situation may occur. *See* Epstein, *Clear View*, *supra* note 21, at 2094–96.

132. Sara's increased opportunity cost is the difference in her opportunity cost without the easement and her opportunity cost with the easement. Thus, it is equal to $300 decreased by $100, equaling $200.

133. The holdup by David is not technically inefficient because the parties as a whole are made better off. While Sara is not as well off as she could have been had David charged only his opportunity cost for the easement, David and Sara together are better off.

since Sara's opportunity cost has risen to $300, David will be able to hold Sara up for some price between $0 and $200.

While some scholars scoff at the beneficiary who holds up the servient estate holder,[134] Professor Carol Rose questions whether the Davids of the world are really the "rascal[s]" they are made out to be.[135] As Professor Rose proclaims, "If we are to take servitudes seriously as property rights, then the neighbor's holdout is perfectly legitimate."[136] In making such an argument, Professor Rose relies on the essence of property rights and protecting the concept of ownership to justify the holdup by the beneficiary.[137] But even Professor Rose places a limit on how far the theory can be pushed. During the drafting of the *Restatement (Third) on Property: Servitudes*, Professor Rose recommended that when entering an easement agreement, parties should be required to "state a limited length of time that they think the [easement] will enhance the development."[138] Once the parties have established the intended lifespan of the easement, then Professor Rose suggests that courts may get involved in the dispute.[139] When the court must get involved to settle disputes ex post and assign rights, the rule being utilized is no longer a property rule but a liability rule.[140] Thus, even Professor Rose, who utilizes property rules to suggest that the beneficiary who holds up the burdened party may not be such a rascal after all, places a limit on how long the law should examine easements through a property lens; once the time period established by the parties has passed, Professor Rose recommends that the easement be viewed under a liability lens.[141]

PART III. THE INEFFICIENCY OF LOOKING THROUGH A QUASI-PROPERTY LENS: THE CIVIL LAW RULE ALLOWING TERMINATION BY NONUSE

By not allowing mere nonuse to terminate easements, the American common law sanctions the inefficient use of easements because it allows for the possibilities of holdups, speculative purchasing, societal loss, and questionable predictions about the future. Unlike the American system,

---

134. *E.g.*, JAMES GORDLEY, FOUNDATIONS OF PRIVATE LAW: PROPERTY, TORT, CONTRACT, UNJUST ENRICHMENT 99–100 (Oxford Univ. Press 2006).

135. Carol M. Rose, *Servitudes, Security, and Assent: Some Comments on Professors French and Reichman*, 55 S. CAL. L. REV. 1403, 1412–14 (1982).

136. *Id.* at 1412.

137. *Id.*

138. *Id.* at 1414.

139. *Id.*

140. *See generally* Epstein, *Clear View*, *supra* note 21 (discussing the distinction between property and liability rules).

141. *See* Rose, *supra* note 135.

other judicial systems, such as the civil law,[142] allow easements to be terminated by nonuse by employing quasi-property rules.

A. The Civil Law Rule Allowing Termination by Nonuse, a Quasi-Property Rule

Servitudes in the civil law are not classified as minutely as they traditionally have been in the common law.[143] Instead, two broad types of servitudes exist: praedial servitudes and personal servitudes.[144] Praedial servitudes are rights *in rem* over a particular estate, whereas personal servitudes are rights *in personam* in a particular beneficiary.[145] Though recognized as a valid form of servitude in the civil law, easements are not separately distinguished in form; they are merely considered part of the broader categories of praedial and personal servitudes.[146]

Praedial and personal servitudes may be terminated after nonuse has occurred for an established time period.[147] This mode of termination—

---

142. Aside from providing a different type of rule for nonused easements, the civil law is a good source to look to for understanding the common law on servitudes given that the English law on servitudes is based heavily on Roman law. *See* OLIVER WENDELL HOLMES, JR., THE COMMON LAW 367 (Dover Publications, Inc. 1991) (1881).

In addition to the civil law, Islamic law also allows easements to be terminated by nonuse. Under the civil law, termination by nonuse only applies to rights relating to real property *other than* ownership, but Islamic law applies the doctrine of extinction by nonuse to ownership. *See generally* SIRAJ SAIT & HILARY LIM, LAND, LAW, AND ISLAM: PROPERTY AND HUMAN RIGHTS IN THE MUSLIM WORLD 12 (Zed Books 2006). Islamic property law is religiously based, with the fundamental sources being the Koran and Sunnah. *See generally* JOHN MAKDISI, ISLAMIC PROPERTY LAW 8 (Carolina Academic Press 2005). Based on these authorities, ownership of land is considered a sacred trust between the landowner and Allah. *See* YAHAYA YANUSA BAMBALE, ACQUISITION AND TRANSFER OF PROPERTY IN ISLAMIC LAW 4 (Malthouse Press Ltd. 2007). Because estate holders are merely serving as God's trustees, Islamic property law emphasizes that land "should be put to continuous productive use;" to do otherwise would be to slight God. SAIT & LIM, *supra*, at 11. As "[l]and ownership in Islam is linked to land use," the nonuse of land can lead to the termination of ownership. SAIT & LIM, *supra*. *See generally* Dan E. Stigall, *A Closer Look at Iraqi Property and Tort Law*, 68 LA. L. REV. 765, 776 (2008).

143. One of the goals of the *Restatement (Third) of Property: Servitudes* was to do away with the numerous classifications of servitudes. Susan F. French, *Tradition and Innovation in the New Restatement of Servitudes: A Report From Midpoint*, 27 CONN. L. REV. 119, 119–20, 124 (1994).

144. *See* BARRY NICHOLAS, AN INTRODUCTION TO ROMAN LAW 140–47 (Clarendon Press 1962).

145. *Id.* at 144. This division is akin to the division of appurtenant and in gross in the common law. *See supra* note 46.

146. *See* NICHOLAS, *supra* note 144.

147. CIVIL CODE [C.C.] art. 942 (Colom.); CODE CIVIL [C. CIV.] art. 706 (Fr.); CODICE CIVIL [C.C.] art. 1073 (Italy); CÓDIGO CIVIL [C.C.] art. 546 (Spain). Though the common law rule against termination by nonuse conceptually only applies to affirmative easements, the

referred to as prescription in the civil law[148]—existed even before the era of Justinian.[149] Prior to Justinian's rise to power, servitudes could be extinguished by two years of nonuse.[150] Justinian's Code amended this rule so that all servitudes were extinguished after nonuse of ten or twenty years.[151] The ten-year rule applied when the parties to the servitude arrangement were present, while the twenty-year rule applied when the parties were absent.[152]

When drafting the Code Napoleon, the French extended the time period provided in Justinian's Code to thirty years,[153] though many civil law jurisdictions today maintain a period of ten[154] or twenty years.[155] The French viewed the perpetuation of Justinian's rule of automatic termination by nonuse as a "just" result because the holder of the dominant estate had the unilateral ability to interrupt the running of the prescriptive period.[156] An interruption occurred by either the beneficiary using the easement or obtaining an acknowledgement of his right in the easement from either a court or the servient estate holder.[157]

Looking at nonused easements in the civil law through the framework of Calabresi and Melamed, the rule that nonused easements are automatically terminated after a certain period of nonuse might be considered a quasi-property rule.[158] The automatic termination of easements by nonuse is like a property rule in that whether the easement ends is in the

---

civil law allows termination by nonuse of what the common law would term negative or spurious easements. Under the civil law, if action inapposite to the "negative" or "spurious" servitude takes place for a statutorily determined period of time, then the servitude is automatically removed from the underlying estate. *See, e.g.*, LA. CIV. CODE ANN. art. 721 (2008).

148. G. BAUDRY-LACANTINERIE & A. TISSIER, TRAITÉ THÉORIQUE ET PRATIQUE DE DROIT CIVIL, PRESCRIPTION (4th ed. 1924) nos. 586, 592 *reprinted in* 50 REVUE TRIMESTRIELLE DE DROIT CIVIL 287, 292 (La. State Law Inst. trans., West 1972).

149. CODE JUST. 3.34.13–14 (Justinian 531). Justinian "restor[ed] the glory of the Roman Empire" when he became Emperor in 527. NICHOLAS, *supra* note 144, at 14. During his reign, he codified the law in what became known as the *Corpus Iuris Civilis*. *Id.* at 39.

150. CODE JUST. 3.34.13–14 (Justinian 531). The rule that nonuse can terminate rights in real property has existed as far back as Hammurabi's Code. Under Hammurabi, if a lessee failed to use the leased land for three years, the lessee had to return the land to the lessor. CODE HAMMURABI ¶ 44.

151. CODE JUST. 3.34.14 (Justinian 531).

152. *Id.*

153. C. CIV. art. 706 (Fr.).

154. LA. CIV. CODE ANN. arts. 622 (1980), 753 (2008).

155. *See e.g.*, C.C. art. 546 (Spain); C.C. art. 1073 (Italy).

156. *See* 1 MARCEL PLANIOL, TREATISE ON THE CIVIL LAW, 758 (La. State Law Inst. trans., 12th ed. 1959) (1939).

157. *Id.*

158. *See The Cathedral, supra* note 1.

hands of the holder of the easement, the beneficiary.[159] As long as the dominant estate holder uses the easement, it will continue to exist. If the arrangement between David and Sara was that David could drive across Chateau Noir, then as long as David uses the easement on the last day of the twenty-ninth year, his easement would remain intact for at least the next thirty years.[160] Prior to the thirty-year mark, however, the easement cannot be terminated without David's consent.[161] Thus, the civilian rule is like a property rule.[162]

Where the rule differs from a traditional property rule, though, is that once thirty years pass, the servitude will be terminated without David's consent if he has not used it.[163] That the easement may terminate without David's consent is a liability-like feature of the civil law rule.[164] However, to consider the civil law rule a liability rule would be incorrect given that the nonholder of the servitude, Sara, cannot terminate the servitude at her whim—the termination will only occur based on the inaction of David.[165] Moreover, Sara need not compensate David for the extinguishment: the servitude simply vanishes as a matter of law.[166] Thus, the civil law rule is a

---

159. *See* RESTATEMENT (THIRD) OF PROP.: MODIFICATION AND TERMINATION OF SERVITUDES § 7.4 cmt. b (2000) ("Abandonment is normally used to describe a situation in which a servitude has terminated because all beneficiaries have relinquished their rights to use an easement.").

160. C. CIV. art. 706 (Fr.). In this case, David's use interrupts prescription because it is use in the manner contemplated by the easement arrangement. In civil law jurisdictions, a use sufficient to interrupt prescription must be a use "in the manner contemplated by the grant or reservation." Goldsmith v. McCoy, 182 So. 519, 523 (La. 1938) (quoting Louisiana Petroleum Co. v. Broussard, 135 So. 1, 2–3 (La. 1931)); *see also* Ashland Oil Co. v. Palo Alto, Inc., 615 So. 2d 971, 973–74 (La. Ct. App. 1991) ("[*Goldsmith*] applied where the manner of use was inconsistent with a broadly or generally worded grant of a servitude.") Thus, prescription running against an easement granted for the purpose of serving as a public road is not interrupted by the dominant estate holder merely driving across the easement; the easement must be developed into a public right-of-way. Southern Amusement Co. v. Pat's of Henderson Seafood & Steak, Inc., 871 So. 2d 630, 639 (La. Ct. App. 2004).

161. C. CIV. art. 706 (Fr.).

162. *Compare* RESTATEMENT (THIRD) OF PROP.: MODIFICATION AND TERMINATION OF SERVITUDES § 7.4 cmt. b (2000) (exploring that abandonment occurs when the beneficiary relinquishes his/her rights), *and* C. CIV. Art. 706 (Fr.) ("A servitude is extinguished by a non-user during thirty years.").

163. *Compare* United Nat. Gas Co. v. James Bros. Lumber Co., 191 A. 12, 14 (Pa. 1937) ("Mere nonuser [sic] does not constitute abandonment; there must be an intention to abandon, together with 'external' acts by which such intention is carried into effect. . . ."), *and* C. CIV. Art. 706 (Fr.) (terminating easements after a set period of nonuse).

164. *See The Cathedral*, *supra* note 1, at 1110.

165. *See* RESTATEMENT (THIRD) OF PROP.: MODIFICATION AND TERMINATION OF SERVITUDES § 7.4 cmt. b (2000)

166. *See* Presault v. United States, 100 F.3d 1525, 1545 (Fed. Cir. 1996) (court recognizes then the owner abandons the easement, the servient estate holder is automatically "relieved").

property rule with a few liability rule features, or more simply put, the rule is a quasi-property rule.[167]

B. The Inefficiency in Applying a Quasi-Property Rule

The inefficiencies of the American common law against termination by nonuse are caused by the application of a property rule to nonused easements.[168] Given that the civil law does not view easements under a pure property lens, but instead looks through a quasi-property lens, some of the problems created in the common law are remedied. One noted problem of the common law is that because it does not allow the servient estate holder to easily alter the arrangement, the burdened party is essentially required to predict his future opportunity costs as well as the opportunity costs of his successors.[169] Because parties are unable to forecast their long term opportunity costs, the common law creates the possibility of long term loss for the servient estate holder. This is not as great of a problem under the civil law because the law automatically removes the easement after a certain period of nonuse.[170] Thus, there is a temporal restriction on the amount of loss that the servient estate holder may accumulate.[171] A similar limitation is placed on the amount of societal loss that can be suffered under the civil law rule.[172] Whereas the American rule allows for these gains to go perpetually unrecognized,[173] the civil law rule provides a cap on how long the societal loss may continue.[174]

Because the quasi-property rule of the civil law does not allow nonuse to continue indefinitely, long-term speculative purchasing on the part of the dominant estate holder is also discouraged.[175] Beneficiaries, like David, may forecast that Blackacre will eventually transform into Eden, but David must act on that speculation within an established time frame. The civil law rule is not a perfect remedy in this regard because David does have the ability to interrupt the prescriptive period by merely using the easement, but the civil law rule may lessen speculative purchasing to some extent.

---

167. *See The Cathedral*, *supra* note 1.

168. *See* Part II.B.

169. *See* Part II.B.1.

170. *See* Part III.A.

171. C.C. art. 942 (Colom.); C. CIV. art. 706 (Fr.); C.C. art. 1073 (Italy); C.C. art. 546 (Spain).

172. *See* Part II.B.2 (explanation of the societal loss suffered under the common law rule against termination by nonuse).

173. *See* Crystal Farms, Inc. v. Road Atlanta, L.L.C., 690 S.E.2d 666, 669 (Ga. Ct. App. 2010) (acknowledging that at common law mere nonuse, without other intentional acts to abandon, is not abandonment).

174. *See* C.C. art. 942 (Colom.); C. CIV. art. 706 (Fr.); C.C. art. 1073 (Italy); C.C. art. 546 (Spain).

175. *See* Part II.B.3 (explaining how speculative purchasing is allowed under the common law rule).

The civil law rule also provides at least some (albeit minimal) relief for the holdup problem that arises under the common law.[176] Because the civil law places a time limit on how long an easement can remain unused, the dominant estate holder is restricted in how long he may hold out in negotiations with the holder of the servient estate.[177] This is a less-than-perfect solution for the holdup problem, however, because the dominant estate holder has the option of using the easement and restarting the prescriptive period, thereby restarting the period in which he may holdup the servient estate holder.[178]

Thus, looking at nonused easements through a quasi-property lens remedies some of the inefficiencies created by the common law property rule; however, the quasi-property lens also creates its own inefficiencies. These inefficiencies stem from the inability of a governing body to predict parties' optimal level of use and the promotion of the overuse or underuse of easements.[179]

1. Inability to Predict Optimal Use

By providing a minimum amount of time within which an easement must be used, the civil law sets a floor for the quantity of use the dominant estate holder must make of the easement.[180] Put another way, civilian systems have pre-determined the maximum amount of nonuse for which parties may contract in an easement arrangement. While this time period may be shortened by the parties,[181] under traditional civil law, it may not be lengthened.[182]

---

176. *See* Part II.B.4 (explaining how the common law rule allows for the dominant estate holder to holdup the servient estate holder).

177. *See* BAUDRY-LACANTINERIE & TISSIER, *supra* note 149.

178. *See id.*

179. *See* Bell and Parchomovsky, *supra* note 17.

180. C.C.art. 942 (Colom.); C. CIV. art. 706 (Fr.); C.C. art. 1073 (Italy); C.C. art. 546 (Spain).

181. *See* Bodcaw Lumber Co. v. Magnolia Petroleum Co., 120 So. 389 (La. 1929); Le Bleu v. Le Bleu, 206 So. 2d 551, 554 (La. Ct. App. 1967); Ober v. McGinty, 66 So. 2d 385, 386 (La. Ct. App. 1953); BAUDRY-LACANTINERIE & TISSIER, *supra* note 148.

182. Louisiana Health Serv. & Indem. Co. v. McNamara, 561 So. 2d 712, 719 (La. 1990); BAUDRY-LACANTINERIE & TISSIER, *supra* note 148. Under traditional civil law, contractual freedom as it applied to prescription was quite limited. For example, under the Greek Civil Code, all transactions amending the conditions of prescription are null and void. Astikos Kodikas [A.K.] [Civil Code] art. 275 (Greece). Under French civil law, prescription could be shortened but not lengthened. *E.g.*, C. Civ. Dec. 4, 1895, D. 96. 1. 241 (Fr.); Req. Nov. 15, 1909, S. 1911. 1. 253 (Fr.). However, there have been recent changes to these limitations on parties' contractual freedom. The recent *Réforme de la prescription en matière civile* revised French Civil Code article 2254 such that it now allows parties to increase or decrease the applicable prescription period. *See* Law No. 2008-561 of June 19, 2008, Gaz. Pal., 2008, 2, 2487.

The problem with setting a use floor—or, viewed from the opposite perspective, a nonuse ceiling—is that just as individual contracting parties are unable to predict their eternal opportunity costs,[183] the government is likewise unable to predict the optimal amount of use and nonuse for a particular easement. In fact, the government is in a worse position to make forecasts regarding the use or nonuse of any particular easement because the government has no information about each individual easement or each particular dominant and servient estate holder.[184]

Moreover, the approach civilian jurisdictions have taken in establishing prescriptive periods for nonused easements is a shotgun approach: civil law systems create one time period governing all types of easements, and even more broadly, all forms of servitudes. The problem with this approach is that a variety of easement arrangements exist, and the maximum level of nonuse (or minimum level of use), is not necessarily the same for each easement situation.[185] David may contract for the right to drive across Blackacre, or he may contract for the right to lay a pipeline across the land. The optimal amount of nonuse for a pipeline arrangement between David and Sara is likely not the same as the optimal amount of nonuse for a right-of-way between the parties. Regardless of what the most efficient time limitation is in each situation, David and Sara are in a much better position than the government is to choose that limitation.

2. Allowing Underuse or Encouraging Overuse

Assuming that the government is incapable of accurately predicting the optimal level of nonuse of an easement, there are two options for the time period a civilian system chooses: either the government can choose an amount less than the optimal level of use or the government can choose an amount greater than the optimal level of use.[186] If the government picks the former, then the level of use chosen will allow for the underuse of easements.[187] Allowing for the underuse of easements gives rise to the societal loss problem that exists in the common law.[188] While an easement is underused, the servient estate holder could be utilizing the estate in a more efficient manner that might create societal gains.

---

183. *See supra* Part II.B.1.

184. *See generally* Bell and Parchomovsky, *supra* note 17.

185. *See* C. CIV art. 707 (Fr.) ("The thirty years begin to run, according to the different kinds of servitudes, either from the day when one ceased to enjoy them, with respect to discontinuous services, or from the day when an act contrary to the servitude has been performed, with respect to continuous servitudes.").

186. *See* BAUDRY-LACANTINERIE & TISSIER, *supra* note 148.

187. *See generally* Daphna Lewinsohn-Zamir, *More is Not Always Better than Less: An Exploration in Property Law*, 92 MINN. L.REV. 635 (2008) (discussing the implications of underusing property).

188. *See supra* Part II.B.2.

If, on the other hand, the required amount of use is set too high, i.e. the time period in which the beneficiary must use is too short, overuse of easements will result.[189] In order to maintain the easement, the beneficiary will be required to use the easement more than is optimal. This encourages a wasteful use of the easement over Blackacre by David. In other areas of property law, however, doctrines are created to discourage or prohibit the wasteful use of property.[190] By setting a strict limit on the amount of nonuse, the civil law of easements promotes the overuse and wasteful use of property.

Like the common law property rule, the civil law quasi-property rule is far from perfect. Underuse, overuse, and predictions of optimal use are all side effects of viewing nonused easements through the civilian quasi-property lens. Recognizing the inefficiencies created in both systems, however, may allow for the deduction of a more efficient solution.

PART IV. THE BELL AND PARCHOMOVSKY GLOSS

The automatic application of static property or quasi-property rules to nonused easements has the potential to create inefficient results. That said, viewing easements through a property lens is not without some merit because property rules help encourage investment and development.[191] To adequately encourage development, the holder of the easement must be able to reap all of the benefits of the land.[192] If the easement holder cannot gain all of the benefits of the easement, he is less inclined to invest in the easement because he will not receive the full payoff for his investment. Property rules prevent easements from being terminated at the whim of the servient estate holder.[193] If the servient estate holder was able to unilaterally end an easement, the easement holder might be less inclined to develop the easement because she might not recognize all of the gains from her investment.

---

189. *But see* Halsrud v. Brodale, 72 N.W. 2d 94, 98 (Iowa 1955) (discussing that the beneficiary cannot increase or extend the use of the easement that was not included in the grant.).

190. Examples of doctrines designed to prevent the wasteful use of property include the rule of capture and the doctrine of waste. *See* Howell v. Union Producing Co., 392 F.2d 95, 98–99 (5th Cir. 1968) (rule of capture); POSNER, *supra* note 16, at 73–74 (doctrine of waste).

191. *See* Bell and Parchomovsky, *supra* note 17, at 27.

192. *See* Harold Demsetz, *Toward a Theory of Property Rights*, 57 AM. ECON. REV. 347, 350–54 (1967) (arguing that property rights internalize costs and benefits thus creating economic gains); Adam Mossoff, *What is Property? Putting the Pieces Back Together*, 45 ARIZ. L. REV. 371, 402 (2003) ("The existence of property therefore is a natural and logical part of the development of human society: it internalizes costs and benefits and thus better effectuates economic activity.").

193. *See generally* RESTATEMENT (THIRD) OF PROP.: MODIFICATION AND TERMINATION OF SERVITUDES § 7.4 cmt. b (2000).

Investment and development is an excellent, if not the best, reason to apply property standards to rights in real property.[194] When an easement becomes nonused, however, investment and development automatically cease. A nonused easement is by definition not in use, and thus not being developed.[195]

Therefore, it appears that the reason to apply property standards to easements vanishes upon nonuse, and only the aforementioned problems remain. If that is correct, in order to continuously employ the most efficient entitlement protection to easements at the point in which an easement becomes a nonused easement, the governing rule should shift from a property standard to a liability rule; nonuse should act as a trigger for a new set of rules.

The notion that property and liability rules may be dynamic was recently discussed by legal scholars Abraham Bell and Gideon Parchomovsky.[196] Building on the Calabresi and Melamed framework, Bell and Parchomovsky proposed that entitlement protections may shift between property and liability rules to create what they referred to as "pliability" rules.[197] The scholars recognize six pliability rules. The first rule is the classic pliability rule, which allows for the transforming of a property rule into a liability rule upon the occurrence of a triggering event.[198] The second rule is the zero order pliability rule, which allows for an initial property rule to morph into a no-liability rule meaning that upon the occurrence of the triggering event, no third party automatically gains a superior right to the original entitlement holder; instead, there is simply open, common access over the formerly protected entitlement.[199] The third rule is the simultaneous pliability rule. Under this rule, an entitlement is governed by both liability and property rules, but there is no discrete triggering event that delineates when governance by one rule ends and the other begins; rather, what rule governs at any moment depends upon the type of use of the entitlement.[200] The fourth rule is the loperty rule, which operates in the reverse of the classic pliability rule, meaning that the entitlement protection initially employed is a liability rule, and upon the occurrence of a triggering

194. *See* Rose, *supra* note 21.

195. Nonuse, in this context, means complete and total nonuse. The concept of complete and total nonuse inherently carries with it an element of time, i.e. nonuse must continue over some period of time.

196. *See* Bell and Parchomovsky, *supra* note 17. To their credit, Bell and Parchomovsky note that other scholars, and perhaps even Calabresi and Melamed themselves, have flirted with the concept of dynamically applying property standards and liability rules. *See id.* at 25 n.102.

197. *Id.* at 5.

198. *Id.* at 31–32.

199. *Id.* at 39.

200. *Id.* at 49–50.

event, a property standard kicks in.[201] The fifth rule is the title shifting pliability rule. This rule allows for the transfer of "property rule protections from one entitlement holder to another."[202] Finally, the sixth rule is the multiple stage pliability rule, which does not restrict the number of shifts from property to liability rules (or vice versa).[203] Instead, under the multiple stage pliability rule, the rule governing an entitlement may change multiple times based on multiple different circumstances.[204]

The six rules described by Bell and Parchomovsky are distinct, but there is a unifying theme among them: they all allow changed circumstances to be incorporated into a legal rule by identifying the change (or changes) and allowing that change to serve as a trigger that shifts protection mode.[205] This idea that a change in circumstance may change the mode of protection is perhaps as simplistically elegant as *The Cathedral*'s original structure.

PART V. THE EFFICIENCY OF LOOKING THROUGH A PLIABILITY LENS: PRIVATE EMINENT DOMAIN BASED ON NONUSE

Applying the basic Bell and Parchomovsky theory that a triggering event may change how an entitlement is governed, nonuse could serve as the triggering event to switch from viewing easements under a property (or quasi-property) lens to looking at easements through a different lens, namely a liability lens.[206] It has already been demonstrated how the benefits of property rules no longer apply once an easement becomes nonused.[207] However, to accurately determine whether nonuse should serve as a trigger to employ liability rules, the benefits of applying liability rules to easements no longer in use must first be identified. If benefits exist such that it is more efficient to apply liability rules to nonused easements as opposed to property standards, then nonuse is a good trigger point under the Bell and Parchomovsky gloss because it will allow the most efficient rule to always be employed: property standards can apply when the easement is in use and the easement holder's rights should be protected, and liability rules can apply when the easement is no longer in use and the non-holder's rights should be afforded greater protection. It must be determined, however, how nonuse would practically act as the trigger point.

---

201. *Id.* at 53–54.
202. *Id.* at 54.
203. *Id.* at 59.
204. *Id.*
205. *See id.* at 67.
206. *See generally The Cathedral*, *supra* note 1.
207. *See supra* Part II.

A. Benefits of Applying Liability Rules to Nonused Easements

Liability rules give the nonholder of the entitlement the opportunity to extinguish the entitlement by paying an objectively determined value to the easement holder.[208] In the context of nonused easements, this means that a liability rule would allow the servient estate holder to end the easement by compensating the dominant estate holder.[209] By looking at nonused easements through a liability lens, it becomes clear that a pure liability rule would prevent the aforementioned inefficiencies of both the common law and the civil law rules. Moreover, application of a liability rule would create additional benefits.

1. Remedy Inefficiencies of the Common and Civil Law

Because liability rules allow for the termination of the agreement by the servient estate holder, an optimal level of nonuse—or at least a level of nonuse closer to the optimal level—will always result.[210] Under liability rules, the parties to the easement decide when a nonused easement terminates, as opposed to the government deciding when an easement should terminate from nonuse. Compared to the government, the individuals in the easement will have superior knowledge about the optimal level of nonuse; thus, the civil law problems created by allowing the government to decide the most efficient amount of nonuse will not arise.[211]

Just as the government is not required to choose the optimal level of nonuse for an easement, the individual parties to an easement arrangement are also not required to predict all of their future opportunity costs under a liability regime. A liability rule would allow for changing conditions to occur; if the servient estate holder mis-forecasts her future opportunity costs, she can extinguish the easement, compensate the dominant estate holder, and then be able to recognize her actual opportunity costs.[212]

A liability rule also discourages societal loss.[213] Once the servient estate holder realizes that she can create greater profit from Blackacre by doing

---

208. *The Cathedral*, *supra* note 1, at 1092; *see* Louis Kaplow and Steven Shavell, *Property Rules versus Liability Rules: An Economic Analysis*, 109 HARV. L. REV. 713, 757 (1996).

209. *See generally id.*

210. Admittedly, this is the optimal level of nonuse in the servient estate holder's eyes. However, based on the manner in which just compensation is determined, the dominant estate holder will also be encouraged to release the servient estate holder at his optimal level of nonuse. See Part V.B.2 for more information regarding just compensation.

211. See Part III.B.1 for a discussion of how the civil law allows the government to predetermine the most "efficient" level of nonuse.

212. There is no concern under the current law of the dominant estate holder misforecasting his future opportunity costs; he can simply release the servient estate holder if he chooses. *See* Part II.B.1.

213. See Part II.B.2 for a discussion of how societal loss is caused by the common law

something other than protecting (or at least not interfering with) a nonused easement, the servient estate holder can terminate the easement. Upon termination of the easement, the servient estate holder is likely to pursue her newfound opportunity, thus creating societal gains.

Since the servient estate holder is able to terminate the easement, liability rules discourage secret, speculative purchasing by the dominant estate holder.[214] If David purchases an easement over Blackacre with the intention of using it only when Blackacre transforms into Eden, he must tell Sara of this plan at the outset of their easement negotiations, or he will risk having the easement taken from him by Sara in return for some objective compensation. Thus, under a liability regime, David is still allowed to gamble, but he must explicitly contract to make that gamble. A liability rule that diminishes speculative purchasing may not seem like a major advancement over the application of a property standard because even when property standards are employed, the common law also allows for parties to contract around speculative purchasing. However, under the common law rule against termination by nonuse, because the easement automatically runs with the land, the onus for instigating contractual negotiations against speculative purchasing is on the servient estate holder. This is a faulty default rule because the dominant estate holder is in a better position to know whether he will be engaging in speculative purchasing as he is the speculative purchaser. As seller to the speculative purchaser, Sara has no way of knowing the underlying rationale for David's behavior. Because David has better information about his own speculative purchase, the burden of contracting for such behavior should be on him. A liability rule, unlike a property standard, places this burden on David. Thus, the shift in the default rule regarding which party bears the burden of instigating contractual negotiations allows for more open, and therefore more efficient, easement transactions.

Finally, holdups are largely prevented by a liability rule. A dominant estate holder cannot holdup a servient estate holder because the servient estate holder has another alternative to transacting with the dominant estate holder: the servient estate holder can obtain a court order that objectively determines the value of the easement. Because the servient estate holder has options for ridding her land of the easement, the dominant estate holder is not in the position to hold her up.

2. Produce Superior Information

In addition to solving the problems created by the common and civil law property and quasi-property rules, liability rules also produce additional benefits. As stated above, liability rules prevent the dominant estate holder from speculatively purchasing the easement without the servient estate

---

rule.

214. *See generally id.*

holder's knowledge. More broadly, using a liability lens to examine nonused easements produces clearer information for both parties throughout the entire easement arrangement.

Initially, liability rules give parties an impetus to negotiate openly with one another and discover the true incentives each party has for entering the easement arrangement. If David wants to purchase the easement as an option-like contract, liability rules demand that he reveal that information to Sara. Sara can better use that information to approximate the cost of the easement to her, and in turn, she may be able to decrease societal loss. If David tells Sara that he really only wants to use the easement once Blackacre magically transforms into Eden, then Sara can contract with David to allow her to use Blackacre for farming wheat until the transformation occurs (if it ever does). In this situation, David receives what he wants—the ability to drive through Eden should it ever appear—and Sara is able to continue using Blackacre productively until the time arrives for David to use the easement. Because Sara is able to use the property while David is awaiting the transformation, she can lower the price of the easement because her opportunity costs diminish. Moreover, societal loss will not occur while David is patiently waiting for his time of use to arrive as Sara is able to utilize the land.

This production of superior information continues throughout the easement arrangement. If suddenly Blackacre becomes prime real estate for the mega shopping complex, Sara can exit the easement arrangement, so long as she pays David just compensation. If David loses faith in the second coming of Eden, then his just compensation would be lower than if he believes the moment of the transformation is fast-approaching. Because David's valuation of the easement is tied to his just compensation,[215] David is encouraged to produce superior information about his own costs throughout the easement arrangement. By having up-to-date information about the opportunity costs of both Sara and David, Blackacre is kept in a constant state of efficiency.

3. Assign Risk of Nonuse to Nonusing Party

In addition to providing superior information, using a liability lens to look at nonused easements also places the risk of losing the easement on the appropriate party. Because a liability rule allows the servient estate holder to terminate the nonused easement, the risk of nonuse falls upon the dominant estate holder. Requiring the dominant estate holder to bear this burden makes sense because the beneficiary is the party in the position to control the nonuse. The party to the arrangement who can control whether use occurs should be the party that takes on the risk of not using the easement.

---

215. Valuation of just compensation is discussed in Part V.B.2.

In sum, liability rules produce great benefits when applied to nonused easements. In addition to solving the problems present in the common law and civil law, applying a liability rule to nonused easements also places the risk of nonuse on the nonusing party and encourages the production of superior information, which, in turn, allows the servient estate to be kept in a constant state of efficiency. Upon this recognition that liability rules are more efficient for governing nonused easements than property standards, it becomes clear that nonuse is an ideal trigger according to the Bell and Parchomovsky pliability theory.[216]

B. Creating a Pliability Regime: Private Eminent Domain Based on Nonuse

That a pliability regime might more efficiently govern easements that become nonused is the easier portion of the analysis; the harder question is how to actually implement this more efficient regime.[217] Under the Bell and Parchomovsky theory, there are six different possibilities for how the pliability regime could be structured.[218] Each form of the pliability rules the scholars promote might be advantageous depending upon, at a minimum, societal goals, the type of easement in question, and the parties involved in the easement arrangement. For the sake of initial analysis as to how a pliability regime might be enacted, it is simplest to utilize the classic pliability rule.

The classic pliability rule allows for the transformation of a property rule into a liability rule upon the occurrence of a triggering event.[219] In the case of nonused easements, that triggering event would be the easement falling into a state of nonuse. Thus, prior to an easement being nonused, i.e. when the easement is a used easement, the classic property rule will apply. Once the easement becomes a nonused easement, liability rules will govern such that the servient estate holder will be allowed to take the easement away from the dominant estate holder by compensating the dominant estate holder with some objectively determined amount.

---

216. *See generally* Bell and Parchomovsky, *supra* note 17.

217. At this point, one might argue that the civil law system should itself be classified as a pliability rule under the Bell and Parchomovsky gloss because it allows for a shift in how the easement is protected. Though whether the dominant estate holder owns the easement does change under the civil law rule, how the easement is protected does not actually change. During the period the civil law allows for nonuse, the dominant estate holder's right to the easement is protected by a property standard. Once the period of nonuse has run, the easement automatically reverts back in ownership to the servient estate holder. Thus, there is no longer an easement; the former easement is extinguished by confusion. As such, though the dominant estate holder's rights shift in the civil law, there is never a second means employed to govern the easement because the easement vanishes.

218. *See generally* Bell and Parchomovsky, *supra* note 17.

219. *See id.* at 29.

This description of a pliability rule for easements is akin to the most recognizable liability rule: eminent domain. The power of the government to seize private property and convert it to public use is the archetype of a liability rule.[220] Under eminent domain, the government is able to unilaterally take the landowners' real property without the landowner's consent, but is required to pay just compensation.[221] By using eminent domain as the foundation for the rule, and using nonuse as the trigger point for enacting this private form of eminent domain, a system of private eminent domain based on nonuse is developed. Implementation of the pliability rule of private eminent domain based on nonuse requires two determinations: first, how to implement nonuse as the trigger point, and second, once nonuse occurs, how to compensate the servient estate holder.

1. Setting Nonuse as the Trigger Point

a. Clearing the Initial Constitutional Hurdle

From an efficiency standpoint, nonuse is the appropriate trigger for the aforementioned pliability rule of private eminent domain. However, the notion of allowing the nonconsensual termination of an individual's property right by another private party for her personal benefit may send up red flags of an unconstitutional taking.[222] Traditionally, the law has not allowed private parties to take property from other private parties for the formers' sole benefit.[223] This principle prevented a group of farmers from

---

220. For a definition of eminent domain, see BLACK'S LAW DICTIONARY 562 (8th ed. 2004). Calabresi and Melamed use eminent domain as the example of a liability rule. *See The Cathedral*, *supra* note 1, at 1106–07. Bell and Parchomovsky use eminent domain as an example of a multiple stage pliability rule. *See* Bell and Parchomovsky, *supra* note 17, at 59–64. While Bell and Parchomovsky refer to eminent domain as a multiple stage pliability rule, it is used here as the basis for a classic pliability rule. The Bell and Parchomovsky analysis finds the doctrine to fall into the sixth category of pliability rules because they view the eminent domain as having three steps: first, an individual's property is protected by a property standard; second, the individual's property is subject to a liability rule such that the Government may take the property; third, the now Government's property is subject to a property standard. In the case of the easement, there is no need to include the third step because the easement will be destroyed by confusion once the servient estate holder takes the easement. Thus, in the case of easements, the form of eminent domain utilized is a classic pliability rule.

221. 26 AM. JUR. 2D *Eminent Domain* § 2 (2010).

222. Though neither the servient estate holder nor the dominant estate holder might be a government entity, the Takings Clause issue must still be raised in order for the servient estate to take the easement via a liability rule, court action will be necessary. Thus, there is a state actor involved and if the taking is unconstitutional, the rule of *Shelley v. Kraemer* restricting state courts from implementing constitutionally violating actions should apply. *See* Shelley v. Kraemer, 334 U.S. 1 (1948).

223. As the majority writes in *Kelo*, "[I]t has long been accepted that the sovereign may

taking property from a railway company to build a grain elevator in *Missouri Pacific Railway Co. v. Nebraska*.[224] In *Missouri Pacific Railway Co.*, the Nebraska State Supreme Court ordered a railway company to give the farmers the desired land for the grain elevator, but the United States Supreme Court reversed, stating that "[t]he taking by a State of the private property of one person or corporation, without the owner's consent, for the private use of another, is not due process of law, and is a violation of the fourteenth article of amendment of the constitution of the United States."[225]

While *Missouri Pacific Railway Co.* seems to be an insurmountable hurdle for private eminent domain based on nonuse to climb,[226] the landmark case *Kelo v. City of New London* may help remove the barrier.[227] The *Kelo* Court held that economic development was included in the meaning of public use for the purposes of taking private land through eminent domain.[228] In *Kelo*, the City of New London, acting on behalf of a private corporation, was permitted to take real property from individuals as part of executing a comprehensive redevelopment plan approved by the city.[229] In reaching its decision, the Court rejected the argument of the property owners that economic development should not qualify as public use.[230] Instead, the Court stated that "[p]romoting economic development is a traditional and long-accepted function of government. . . . [T]here is no basis for exempting economic development from our traditionally broad understanding of public purpose."[231]

Thus, *Kelo* provides that so long as economic development projects impact the public, such projects are within the meaning of the Fifth Amendment.[232] While converting an individual nonused easement to return to the servient estate appears to not have the same societal impact as the economic development project in *Kelo*, there is still some public impact, and in the aggregate, the public impact could be substantial.[233] In fact, it is

---

not take the property of *A* for the sole purpose of transferring it to another private party *B*, even though *A* is paid just compensation." Kelo v. City of New London, 545 U.S. 469, 477 (2005).

224. Missouri Pac. Ry. Co. v. Nebraska, 164 U.S. 403, 411–12 (1896).

225. *Id.* at 417.

226. The continued citation of *Missouri Pacific Railway Co.* only increases the height of the hurdle. *See Kelo*, 545 U.S. at 477–78; Hawaii Hous. Auth. v. Midkiff, 467 U.S. 229, 241 (1984); Thompson v. Consol. Gas Util. Corp., 300 U.S. 55, 79–80 (1937).

227. *See generally Kelo*, 545 U.S. 469.

228. *See id.* at 484–85.

229. *Id.* at 473–74.

230. *Id.* at 484.

231. *Id.* at 485.

232. *See id.* at 477–80.

233. *See* Part II.B.2 for a discussion on societal loss from nonused easements. Moreover, allowing private parties to take nonused easements by paying just compensation is similar to allowing parties to take nonused easements through adverse use. The only difference is that whereas the adverse possessor of the easement must take affirmative steps

arguable that the promotion of economic development rationale applies even more in the case of nonused easements. In *Kelo*, the holders of the property right used their property rights and received a benefit from those property rights: they lived in their homes.[234] While the private corporation in *Kelo* was theoretically going to create benefits by their redevelopment plan,[235] to determine if the redevelopment plan was efficient, any benefits created by the private corporation had to be offset by the loss of benefits by the individual property owners. In the case of nonused easements, the dominant estate holder is not actually using the easement, thus there is nothing to offset the potential benefits the servient estate holder could reap.

b. Determining the Point of Nonuse

Presuming that the taking of a nonused easement meets constitutional muster, the task must turn to determining when nonuse reaches a point such that the servient estate holder should be allowed to take the easement. As the impetus for enacting the pliability rule of private eminent domain is the promotion of continuous efficiency, the taking should only be allowed at a point when nonuse becomes inefficient. In other words, if taking the easement from the servient estate holder is more efficient than not taking the easement, then the taking should be allowed.

As efficiency is the goal, the taking of the easement should be a Pareto-superior transaction. A Pareto-superior transaction is one in which at least one person is better off and no one is worse off.[236] For nonused easements, Pareto efficiency occurs so long as the value of the land to the servient estate holder with no easement (SE_{NE}) minus the value of the easement to the dominant estate holder (DE_E) is greater than the value of the land to the servient estate holder with the easement (SE_E). Thus, so long as

---

to use the easement, the taker under a theory of private eminent domain based on nonuse must pay the dominant estate holder for the easement.

234. *See Kelo*, 545 U.S. 469.

235. The jury is still out as to whether Pfizer, the private company in *Kelo*, actually created any long-term economic benefits; in late November 2009, Pfizer moved out of the City of New London. Patrick McGeehan, *Pfizer to Leave City That Won Land-Use Suit*, N.Y. TIMES, Nov. 13, 2009, at A1.

236. POSNER, *supra* note 16, at 12. Generally, Pareto efficiency is not used; instead, Kaldor-Hicks efficiency is the traditional form of efficiency used to determine whether a law or policy is actually efficient. *See id.* at 13–14. Kaldor-Hicks efficiency states that outcome is efficient if it creates a result in which the winner could fully compensate the loser while still retaining a surplus, though the winner need not actually compensate the loser. *See id.* In the case of private eminent domain based on nonuse, however, there is no need to apply Kaldor-Hicks efficiency because the servient estate holder will actually be compensated by the dominant estate holder. *See* Part V.B.2. As the dominant estate holder will be compensated, he is not any worse off without the easement, so the shift is Pareto efficient. *See id.* at 13–14.

$$SE_{NE} - DE_E > SE_E$$

then termination of the easement is efficient. Put another way, so long as the servient estate holder values getting rid of the easement more than dominant estate holder values keeping the easement, then termination of the easement is efficient. Mathematically, this means that if

$$SE_{NE} - SE_E > DE_E$$

termination of the easement by the dominant estate holder is efficient. This situation is efficient because the gains made by the servient estate holder in terminating the easement are large enough that the servient estate holder can compensate the dominant estate holder for his loss while still reaping a benefit.[237]

For example, suppose that David stops using his easement to drive across Blackacre. He values his nonused easement at $10. Sara values Blackacre with the easement at $100,000 and without the easement at $100,001. In such a situation, it would be inefficient to terminate the easement because, even though David is not using the easement, he values it more than Sara values getting rid of it. If, on the other hand, David values his easement at $10, Sara values her land with the easement at $100,000, and she values Blackacre without the easement at $200,000, then it is efficient to remove the easement from Blackacre.[238]

At this point, one might question why the law must be involved if the servient estate holder values ridding her land of the easement more than the dominant estate holder values maintaining his easement; if the solution is efficient, arguably the free market would have already dictated that outcome. Theoretically this is true, but recall that in practice, the dominant estate holder will hold up the servient estate holder for an amount higher than what the free market would demand.[239] Thus, the reason the law must be involved through a pliability rule is to determine the servient estate holder's and dominant estate holder's value of the easement, SE_E and DE_E respectfully.[240] As the example illustrates, determining when to terminate based on nonuse requires a factual finding of how much each party values the easement. Inarguably, such a finding will be difficult. The values of

---

237. *See* Posner, *supra* note 16, at 12.

238. *See supra* note 130 for a comparison to the contract theory of efficient breach.

239. *See supra* Part II.B.4.

240. In *The Cathedral*, Calabresi and Melamed stated that if the market valuation of the entitlement is inefficient or unavailable, liability rules are the superior protective method. *The Cathedral*, *supra* note 1, at 1110. Here, there is no traditional market value for SE_E and DE_E, and the value assigned by the parties themselves will be inefficient given the holdup problem. Thus, using the equation $SE_{NE} - SE_E > DE_E$ to determine when nonuse occurs for the purpose of private eminent domain is in line with the arguments put forth in *The Cathedral*.

SE_{NE} and SE_E may be determined by the market price for selling Blackacre with and without the easement. While the numbers would be somewhat subjective, an appraisal of the property could be done with the aid of experts.

Determining the value of the nonused easement to the dominant estate holder, DE_E, is far more challenging because there is no market for easements such that an appraiser could easily develop an objective value for the nonused easement. One initial concern in even attempting to give DE_E a value is that the holdup problem could return: David might artificially inflate how much he values the easement, thereby effectively preventing Sara from ridding Blackacre of the easement by this system of private eminent domain based on nonuse. This concern, however, is minimized by regulating the moment at which David's value of the nonused easement should be used to determine whether ridding Blackacre of the easement is efficient. The value of the nonused easement should be determined based on how much the dominant estate holder valued the easement the day *before* he knew the servient estate holder wanted to rid the servient estate of the easement.[241] If the value of the easement to the beneficiary is taken the day after the dominant estate holder knows that the servient estate holder wishes to rid Blackacre of the easement, then the value of the easement to the dominant estate holder will rise exponentially because there is an increase in demand for the easement. Thus, the value of the dominant estate holder must be calculated the day before he became aware of the desires of the servient estate holder.

Putting the time at which DE_E should be determined aside, the issue remains of how to set an objective value for the nonused easements given the lack of a market. One possible way to accurately calculate the value of the nonused easement is through evidence that establishes how much the dominant estate holder originally valued the easement, how much the dominant estate holder originally intended to use the easement, and how much the dominant estate holder has used the easement. For example, suppose that when David purchased the easement for $100, he intended to drive across Blackacre once a week. After fifty weeks of driving across Blackacre each week, David stops. At this point, the value of the easement to David has potentially decreased. He has had fifty-one weeks of possible use, but only used the easement for fifty of those weeks. Thus, his value of the easement may be calculated as:

$$(50 / 51) \times \$100 = \$98.04$$

Now suppose that David continues to not use the easement for 100 weeks. At this point, he has used the easement for fifty weeks and not used it for

241. The idea of valuing property the day before a demand-altering event occurs is used for dissenters' rights in merger agreements. *See, e.g.*, N.Y. BUS. CORP. § 1118(b) (Consol. 2003).

the following 100 weeks. In this case, the value of the easement to him might be calculated as:

$$(50 / 150) \times \$100 = \$33.33$$

The more that David fails to use the easement, the less he objectively values the easement.

The evidence required to make such evaluations is ascertainable. David's original valuation of the easement can be determined based on the contract price for the easement, and his actual use of the easement can be presented through his own testimony with the corroborating testimony of others who saw him use the easement. The difficult number to obtain will be how much David originally intended to use the easement. Such information could be gained through David's testimony and through Sara's testimony to the extent that David revealed his plans to Sara.

Valuing the nonused easement by relying heavily on David's testimony raises some concerns with the efficiency of instituting this idea of private eminent domain based on nonuse. Law and economics scholars James Krier and Stewart Schwab argue that when valuation by a court is too subjective, such rules are problematic.[242] According to Krier and Schwab:

> Judges will have problems assessing the correct values for the same reason private bargainers would: limited, hidden information. If parties can hide their valuations from each other, they can hide them from a judge. Judges can probably assess subjective values accurately enough when the relevant information is out in the open, but in such cases bargaining might work just as well, since it's hard to be open and at the same time strategic.[243]

In their analysis, Krier and Schwab limit their critique of liability rules to multi-party situations,[244] but their point has value in the bilateral monopoly situation that arises for easements: having the information out in the open would increase the ability of a court to establish a correct value. The benefit of having private eminent domain based on nonuse is that it encourages parties to provide superior information to one another. Thus, David will be encouraged to publically indicate his intended use of the easement. If David intends to use the easement for the first fifty weeks, then head to France for two years, and then return to his tract of land to use the easement, this type of pliability rule encourages him to include such information in the easement agreement. By having such information in the agreement, courts

---

242. *See* James E. Krier & Stewart J. Schwab, *Property Rules and Liability Rules: The Cathedral in Another Light*, 70 N.Y.U. L. REV. 440, 461–62 (1995).

243. *Id.* at 462.

244. *Id.* at 461.

are in a better position to calculate the actual value of the nonused easement to the dominant estate holder.

Given that the valuation of the easement to the dominant estate holder is based in part on the amount the beneficiary uses the easement, a concern might arise that such a rule would encourage overuse of the easement. If David knows that by not using the easement he may lose it, then, much like when the civil law overestimates the optimal amount of use,[245] David will be inclined to use it unnecessarily. While that is certainly a concern, it is less likely to occur under this pliability regime than it does in the civil law's quasi-property regime as parties under the proposed pliability system can contract around the rule. If David wants to not use the easement for an extended period of time, he will be encouraged to include that information in the original agreement. Obviously this places the onus on David to contract for his projected nonuse, but as previously stated, this shift in who bears the burden of predicting nonuse is the more efficient solution given that David is far better equipped to predict his own nonuse than Sara.

In order to rid Blackacre of the easement, the servient estate holder will have to sue the dominant estate holder, so the concern may arise that a system of private eminent domain based on nonuse will increase litigation. Professors Ian Ayres and Paul M. Goldbart, however, argue that liability rules do not actually increase litigation. Professors Ayres and Goldbart state that "[u]nder a liability regime, litigation costs give the parties an additional impetus to negotiate and hence can make liability rules more efficient than property rules."[246] The knowledge that there may be future litigation increases the parties' production of information from the beginning, thus resulting in contracts better-suited for the individual desires of the parties. The superior information included in the contract leads to a lesser chance that the parties will engage in litigation in the long run.

2. Structuring the Compensation

Upon determining the point of nonuse at which private eminent domain may occur, the next question to answer is how to compensate the dominant estate holder. The dominant estate holder should be compensated at least the amount at which he values his current use of the easement. This means the dominant estate holder should receive, at a minimum, the value he will lose from termination of the easement. This allows the dominant estate holder to be compensated for the loss he incurs from termination of the easement, but it does not compensate the dominant estate holder for the amount of the easement that he is not using. Such compensation is unnecessary because when the dominant estate holder failed to use his

245. *See supra* Part III.B.2.

246. Ian Ayres and Paul M. Goldbart, *Optimal Delegation and Decoupling in the Design of Liability Rules*, 100 MICH. L. REV. 1, 61 (2001).

easement, the servient estate holder still had the responsibility of protecting the easement for the dominant estate holder's potential use.

For example, Sara was not able to farm wheat regardless of whether David used the easement, as farming wheat would have impeded David's ability to drive across Blackacre had he opted to use the easement. Thus, David's easement received more protection from Sara than it required. In receiving this extra protection, David was compensated despite his nonuse. If Sara is made to pay David for that period of nonuse, then David is being paid twice.

The secondary benefit of requiring compensation to the dominant estate holder be at least at the amount of the used easement is that it encourages him to negotiate with the servient estate holder sooner rather than later. If Sara tries to take David's easement through private eminent domain based on nonuse, then the longer David has not used the easement, the less compensation he may receive because his value of the nonused easement grows smaller the longer the easement sits in a state of nonuse. Thus, David has an incentive to try and negotiate with Sara as soon as he knows he will no longer use the easement. At this point, he can bargain with Sara to be paid for releasing Blackacre from the burden of the easement for a higher price than he might receive from a court later on.[247]

In addition to compensating the dominant estate holder for the nonused portion of his easement, the law should also recognize that the servient estate holder cannot merely be let off the hook for initially engaging in a bad easement arrangement. When the value of the servient estate increases such that continuing the easement arrangement is inefficient for the servient estate holder, the servient estate holder should have to pay the dominant estate holder some additional amount for being released from the easement arrangement. However, the total amount the servient estate holder pays the dominant estate holder cannot exceed the increased value that the servient estate holder will receive by ridding the burdened estate of the easement. In other words, the total amount paid to the dominant estate holder cannot be greater than the difference between the price of the land without the easement (P_{NE}) and the price of the land with the easement (P_E).

Thus, in total, just compensation must be somewhere between the difference of the price of the land without the easement (P_{NE}) and the price of the land with the easement (P_E), and the price of the nonused easement (P_{NU}). Mathematically, the amount of just compensation (JC) can be expressed as:

$$P_{NE} - P_E > JC > P_{NU}$$

This means that a servient estate holder will never try to rid Blackacre of the easement if the value of the easement to the dominant estate holder exceeds the increased value of the servient estate sans easement. So, when

247. Because the compensation scheme suggested should encourage negotiation, it helps to discourage litigation, a concern addressed in Part V.B.1.

$$P_{NU} > P_{NE} - P_E$$

Sara will not try to remove the easement from Blackacre, even if there is nonuse. Sara will only be incentivized to use private eminent domain based on nonuse when a court will allow an easement to be terminated under the aforementioned inefficient nonuse standard.[248] Under the inefficient nonuse standard, if the value of the nonused easement is greater to the dominant estate holder than the increased value of the easement-less estate to the servient estate holder, then the easement cannot be terminated.

By compensating the dominant estate holder at some amount between P_{NU} and $P_{NE} - P_E$, the servient estate holder will pay less than she would under the rule against termination by nonuse. Under the American rule, the dominant estate holder will hold out for the entire increased value of the servient estate, i.e. the entire value of $P_{NE} - P_E$. Additionally, the proposed system prevents the dominant estate holder from holding up the servient estate holder. In fact, all parties are encouraged to remove the easement from the servient estate as soon as the easement has become inefficient under a system of private eminent domain based on nonuse. David is encouraged to enter into an arrangement with Sara as soon as he has stopped using the easement because that is the time when P_{NU} will be the greatest. This means that the least amount that David might be compensated for the easement is at its highest point; if David waits to reach a deal with Sara one month later, then P_{NU} will be lower, so he may receive less compensation. Similarly, Sara has an incentive to reach a deal with David as soon as the easement is inefficient because the longer Sara waits, the greater the difference in P_{NE} and P_E may become because P_{NE} will (presumably) grow. As the difference in P_{NE} and P_E increases, the amount that Sara might have to pay David will increase.

Of course, providing compensation under the theory of private eminent domain based on nonuse revives the earlier criticism of scholars like Krier and Schwab—that the judiciary may assign speculative values.[249] Certainly determining the amount of compensation due is not an easy task and demonstrates that the system presented herein is not perfect. Be that as it may, the pliability rule of private eminent domain based on nonuse is superior in terms of efficiency to the current rules under the common law and civil law systems. The proposed rule allows for the parties with the most information to establish when the nonused easement should be terminated, thereby accounting for changed conditions to both the servient and dominant estates. Private eminent domain based on nonuse discourages speculative purchasing and prevents holdups by allowing the servient estate holder to unilaterally extinguish the easement. And it decreases societal loss because the nonused easement will cease to exist, thus allowing the servient estate holder to use her estate more productively. All in all, looking at

248. *See* Part V.B.1.

249. *Supra* notes 243–45.

nonused easements through a pliability lens is more efficient than using the property lens of the common law or the quasi-property lens of the civil law.

CONCLUSION

The American common law for centuries has examined easements through a property lens and maintained that nonuse alone does not terminate an easement.[250] The civil law has viewed nonused easements through a quasi-property lens, allowing easements to be terminated upon a statutorily established period of nonuse.[251] In doing so, both systems create inefficiencies that could be remedied by governing easements with a liability rule once easements become unused.[252] The more efficient system is to govern easements by a pliability rule of private eminent domain based on nonuse, which uses nonuse as a trigger point to shift from employing a property standard to employing a liability rule.[253]

As developed herein, private eminent domain based on nonuse is a judicially imposed doctrine that requires a fair amount of fact finding of the parties' valuations for the nonused easement.[254] The judicial fact finding necessary to implement private eminent domain inevitably will give some pause; courts, some will argue, are not the most efficient bodies for determining individuals' valuations of rights in real property.[255] Be that as it may, the inability of the market to efficiently govern nonused easements forces the judiciary to take on such an evaluation role.

If this article is correct in asserting that terminating easements due to their nonuse is a more efficient use of property, why not apply the same, or at least a similar, concept to other nonused real property rights? There are some obvious advantages to applying private eminent domain based on nonuse to easements—there is a clear party to bring the action (the servient estate holder), there is a clear recipient of the easement once it has been terminated (again, the servient estate holder), etc.—but that is not to say that the general concept of pliability rules may not be freshly applied to other antiquated property doctrines. If correctly employed, pliability rules ensure that the entitlement-protection method governing real property rights is always the most efficient entitlement-protection method.

And perhaps that is the broader and more important point of this article: there are an endless number of property doctrines in both the common law and the civil law that have been governed by the same, static entitlement-protection methods since their creation. But as our legal theories regarding entitlement protections grow more sophisticated, it is well worth our time to

250. *See supra* Part II.A.
251. *See supra* Part III.A.
252. *See supra* Part III.A.1.
253. *See supra* Part III.B.1.
254. *See id.*
255. *See id.*

re-examine these longstanding property doctrines to determine if they still produce the most efficient results.

REDEEMING AN EMPTY PROMISE: PROCEDURAL JUSTICE, THE CRIME VICTIMS' RIGHTS ACT, AND THE VICTIM'S RIGHT TO BE REASONABLY PROTECTED FROM THE ACCUSED

MARY MARGARET GIANNINI*

ABSTRACT 47
I. INTRODUCTION 48
II. A DIAGNOSIS: THE EMPTY PROMISE OF THE VICTIMS' RIGHT TO BE REASONABLY PROTECTED FROM THE ACCUSED 52
A. *Rhetoric and Reality* 52
B. *Constitutional Constraints on a Right to Protection* 53
C. *State Tort Limits on the Victim's Protection Right* 58
1. Sovereign Immunity, the Public Duty Doctrine and the Special Duty Doctrine 58
2. A State Case Study: Massachusetts, Promises of Protection, and Tort Limits 62
D. *Taking Away What Was Given: Internal Restrictions Within the CVRA* 68
III. A THEORETICAL REVIEW: THE VICTIM IN THE CRIMINAL PROCESS .. 74
A. *The Rise and Fall of Victim Primacy* 74
B. *The Public Prosecution Model* 76
C. *Victims Respond: The Emergence of the Victims' Rights Movement* 82
D. *Procedural Justice and the Victim* 85
IV. A REMEDY: PROCEDURAL JUSTICE AND THE PROMISE OF PROTECTION 96
V. CONCLUSION 103

ABSTRACT

The federal Crime Victims' Rights Act (CVRA) provides victims with a host of rights, including reasonable protection from the accused.

* Assistant Professor of Law, Florida Coastal School of Law, 2008–present. Indiana University School of Law – Indianapolis, J.D., 2002; State University of New York at Albany, M.L.S., 1993; University of St. Andrews, Scotland, M.A., 1992. I would like to thank Florida Coastal School of Law for the research grant that supported this project, as well as Bjorn Anderson, Victoria Calhoon, Jess Dichter and Andrea Horton for their able research assistance. An earlier version of this paper was presented at the Judge Stephanie K. Seymour Lecture in Law, at University of Tulsa College of Law, Tulsa, OK Spring 2009, and at the National Crime Victim Law Institute's 2010 Annual Meeting in Portland, OR.

However, the current protection language in the CVRA offers very little to victims in the form of a meaningful, substantive, and enforceable right. Constitutional principles, tort immunity concepts, as well as other statutory limits within the CVRA itself, constrain the extent to which victims can rely on or enforce a right to protection. While the victim's CVRA right to protection may currently represent an empty promise, this article asserts that the right can be redeemed if it is interpreted and redefined by procedural justice principles. Many of the other rights granted to victims under the CVRA are naturally grounded in procedural justice theory. When framed in this manner, the CVRA affords victims a meaningful role in the prosecution of the offender, while also providing a tangible process by which to enforce their rights. This article proposes ways to bring the victim's protection right into alignment with procedural justice theory as well as with the other rights granted to victims under the CVRA. Viewing the CVRA's right to protection through procedural justice principles, the CVRA will cease to be an empty promise and instead, can serve crime victims in a meaningful way.

I. INTRODUCTION

The Crime Victims' Rights Act (CVRA) represents one of the most far reaching pieces of federal legislation passed by Congress on behalf of crime victims.[1] Among the many rights granted to victims under the statute is "[t]he right to be reasonably protected from the accused."[2] This language, even when tempered by the word "reasonably," suggests that the federal government has assumed the duty to protect victims from further harm by defendants. Certainly, some of the political rhetoric supporting and leading up to the passage of the CVRA suggested as much. For example, Arizona Senator Jon Kyl, one of the law's primary sponsors, noted:

> Congress' concern for the safety of crime victims is appropriate and just. The United States Supreme Court has recognized that the "primary concern of every government . . . [is] for the safety and indeed the lives of its citizens." In the past, victims have been grievously harmed—even murdered—because courts have been inattentive to their needs while making decisions about pre-trial release of the accused.[3]

---

1. Crime Victims' Rights Act, 18 U.S.C. § 3771 (2006).

2. *Id.* at § 3771(a)(1).

3. John Kyl, Stephen Higgins & Steven J. Twist, *On the Wings of Their Angels: The Scott Campbell, Stephanie Roper, Wendy Preston, Louarna Gillis, and Nila Lynn Crime Victims' Rights Act*, 9 LEWIS & CLARK L. REV. 581, 596 (2005) (quoting United States v. Salerno, 481 U.S. 739, 755 (1987)) (alterations in original). In Congressional testimony for a proposed victims' rights amendment to the U.S. Constitution, which eventually resulted in the passage of the CVRA statute, Senator Kyl argued that providing rights to crime victims was "the least the system owes to those it failed to protect." *A Proposed Constitutional*

California Senator Dianne Feinstein, another primary sponsor of the CVRA, indicated that the victim's protection right was intentionally listed first in the statute in order to emphasize and, as Senator Kyl wrote, "reinforce[] the principle that government's first and foremost obligation to its citizens is to protect them—especially those who have already been victims of a crime."[4]

That a victim might rely on such broad rhetoric supporting a right to protection makes a measure of intuitive sense. An initial review of the preamble of the United States Constitution and Declaration of Independence could lead many to believe that "one of the first duties of any government is to offer adequate physical protection to its constituents."[5] The Declaration of Independence suggests that governments should be formed to protect the citizenry's rights to "Life, Liberty, and the pursuit of Happiness."[6] Likewise, the Constitution notes that our founders joined to create the United States of America in order to "establish Justice, insure domestic Tranquility, provide for the common defense, [and] promote the general Welfare . . ."[7] Hence, it should not be at all surprising that the law is peppered with statements articulating the premise that governments are

---

Amendment to Protect Victims of Crime: Hearing Before the S. Comm. on the Judiciary, 105th Cong. 8 (1997) (statement of Sen. Jon Kyl). *See also infra* note 218 (regarding the events which lead up to Congress' passage of the CVRA).

4. Kyl et al., *supra* note 3, at 595. The CVRA lists a victim's rights in the following order:

> (a) Rights of crime victims. – A crime victim has the following rights:
> (1) The right to be reasonably protected from the accused.
> (2) The right to reasonable, accurate, and timely notice of any public court proceeding, or any parole proceeding, involving the crime or of any release or escape of the accused.
> (3) The right not to be excluded from any such public court proceeding, unless the court, after receiving clear and convincing evidence, determines that testimony by the victim would be materially altered if the victim heard other testimony at that proceeding.
> (4) The right to be reasonably heard at any public proceeding in the district court involving release, plea, sentencing, or any parole proceeding.
> (5) The reasonable right to confer with the attorney for the Government in the case.
> (6) The right to full and timely restitution as provided in law.
> (7) The right to proceedings free from unreasonable delay.
> (8) The right to be treated with fairness and with respect for the victim's dignity and privacy.

18 U.S.C. § 3771(a).

5. Richard L. Aynes, *Constitutional Considerations: Government Responsibility and the Right Not to be a Victim*, 11 PEPP. L. REV. 63, 66 (1984).

6. THE DECLARATION OF INDEPENDENCE para. 2 (U.S. 1776).

7. U.S. CONST., pmbl.

created, in part, to provide protection to their citizens.[8] Nor should it be surprising that victims' rights advocates would invoke and rely on these laudatory principles when claiming it is appropriate to grant victims a right to protection.[9] However, the suggestion that the government has an affirmative and enforceable obligation to protect its citizens from harm is tenuous.

Despite the grand rhetoric that accompanied the passage of the CVRA, the victim's right to protection rests on a precarious foundation that undermines the right's substance and enforceability. However, the inadequacies that currently burden the CVRA's protection right are not insurmountable. Just as a majority of the rights granted to victims under the CVRA are best grounded in procedural justice theory, the victim's right to be reasonably protected from the accused can also be viewed through this prism, thereby giving the right substance, meaning, and enforceability.

Section II of this article is diagnostic. This section contrasts the specific protection language in the CVRA and the rhetoric accompanying its passage with the many legal hurdles that limit the right's scope and a victim's ability to enforce it. The CVRA's protection language, thus far, has not received much attention from scholars nor has it been specifically tested in the courts, but several parallel areas in the law highlight the unreliability of a protection right.[10] Constitutional concepts, as well as standard tort

---

8. United States v. Salerno, 481 U.S. 739, 755 (1987) (noting that a "primary concern of every government [is the] concern for the safety and . . . the lives of its citizens"); Metromedia, Inc. v. City of San Diego, 453 U.S. 490, 507–08 (1981) (referencing the authority of local government to protect citizens' legitimate interests in traffic safety); Piemonte v. United States, 367 U.S. 556, 559 n.2 (1961) (noting that "[t]he Government of course has an obligation to protect its citizens from harm"); St. Louis & S.F. Ry. Co. v. Mathews, 165 U.S. 1, 19 (1897) ("If the state is powerless to protect its citizens from the ravages of fires set out by agencies created by itself, then it fails to meet one of the essentials of a good government. Certainly, it fails in the protection of property.") (quoting Mathews v. St. Louis & S.F. Ry. Co., 24 S.W. 591, 596 (Mo. 1893)); Paul v. Virginia, 75 U.S. 168, 176 (1868) ("It is the duty of all governments to pass all laws which may be necessary to shield and protect its citizens.") *overruled on other grounds by* United States v. South-Eastern Underwriters Ass'n, 322 U.S. 533, 543–53 (1944); 515 Assoc's v. City of Newark, 623 A.2d 1366, 1370 (N.J. 1993) ("Without doubt, local governments bear the burden of providing police protection. . . ."); Ex parte Bushnell, 9 Ohio St. 77, 103 (Ohio 1859) ("On these two principles—allegiance to the state, protection to the citizen—rests not merely all sovereignty, but the very social compact itself."). Even the Transportation Security Administration notes on its website that its mission is to "protect[]the Nation's transportation systems to ensure freedom of movement for people and commerce." *See* http://www.tsa.gov/who_we_are/mission.shtm (last visited Jan. 17, 2011).

9. *See supra* note 8.

10. *See, e.g.*, United States v. Rubin, 558 F. Supp. 2d 411, 419–421 (E.D.N.Y. 2008) (noting limited judicial interpretation of right, and questioning whether the right provides anything to victims); United States v. Turner, 367 F. Supp. 2d 319, 322, 336 (E.D.N.Y. 2005) (providing limited review of the victim's right to be reasonably protected from the accused); Eric Blondel, *Victims' Rights in an Adversary System*, 58 DUKE L.J. 237, 269

analysis, severely undermine any dependence a victim might place on a promise of protection. Moreover, other portions of the CVRA further constrain, if not entirely foreclose, a victim's ability to rely on the statutory grant of protection from the accused. What remains is a largely illusory and empty promise.

In an attempt to redeem the CVRA's unfulfilled promise of protection, Section III is devoted to identifying a theoretic framework upon which victims' rights, and more particularly, the victim's right to be reasonably protected from the accused, can rest. In so doing, this section includes a historical review of the victim's role in the criminal justice system, charting the ebb and flow of the victim's predominance within criminal procedure and the eventual ascendance of the public prosecution model. One result of the public prosecution model was that its utilitarian and retributive roots rendered the victim a silent, if not forgotten person within the criminal justice system.[11] The victims' rights movement has gone a great distance in correcting this oversight, but criminal justice theorists still struggle to identify how the public prosecution model can appropriately make room for the victim.[12] In addressing this conflict, I assert that the social science concept of procedural justice provides the medium to create a befitting space for victims within the criminal justice system.

Section IV examines how most of the rights afforded to victims under the CVRA are already grounded in procedural justice principles. I also note that there are other areas within federal criminal procedure that approach victims' rights from a procedural justice posture and could serve as a model for similarly restructuring the victim's protection right. I then conclude that the victim's right to protection should also be framed within procedural justice principles. I must acknowledge, however, that how I use procedural justice theory to reshape victim's protection right undermines an explicit reading of the right as it currently appears in the CVRA. In anticipation of any such criticism, I suggest two slight amendments to the CVRA's protection and enforcement language. By altering the protection language of the CVRA so that it is better grounded in procedural justice theory, I transform the CVRA's empty promise of protection into a substantive and enforceable right for victims.

(2008); *see also infra* notes 316–322 and accompanying text, regarding the *Turner* case.

11. William F. McDonald, *Towards a Bicentennial Revolution in Criminal Justice: The Return of the Victim*, 13 AM. CRIM. L. REV. 649, 650 (1976).

12. *Id.* at 661–62.

II. A DIAGNOSIS: THE EMPTY PROMISE OF THE VICTIM'S RIGHT TO BE REASONABLY PROTECTED FROM THE ACCUSED

A. Rhetoric and Reality

The rhetoric that accompanied the passage of the victim's CVRA protection right was inspiring because it invoked the notion that governments exist to protect the citizenry.[13] The idea that a core government function is to protect its citizens is not entirely unfounded. The social contract theory of government, as developed by such philosophers as Jean-Jacques Rousseau and John Locke, begins with the general proposition that human beings originally existed in a state of nature, where individuals bore the responsibility to protect their own person and property from the wrongdoings of others.[14] As people joined together in more structured societies and governments, they relinquished certain freedoms such as individual acts of self-help and self-defense in response to criminal acts, in exchange for the mutual protection that came from being part of an organized group.[15]

One core feature of the social contract was the understanding that the citizen's promise to obey the laws of society imposed upon the government a duty "to protect the citizenry and punish violators" of the contract.[16] John Locke further emphasized that the "formation of the government, by the people joining into the contract, is governed by public good and with the consent of the individuals agreeing to be governed."[17] Therefore, the trust established between the government and those who had consented to be governed required that the government exercise its power for the good of society.[18] When the government breached this trust by enacting rules that failed to preserve and protect the property and safety of its citizens, the governed had the right "to dissolve the government and create a new one that would protect their rights and guard their safety."[19] It was upon these very principles that the United States of America was founded.[20]

Despite the underlying idea that governments exist to provide collective protection to its citizens, courts and legislatures have largely rejected any

---

13. *See supra* notes 3–4 and accompanying text.

14. Liliya Abramchayev, *A Social Contract Argument for the State's Duty to Protect from Private Violence*, 18 ST. JOHN'S J. LEGAL COMMENT. 849, 849–50 (2004).

15. *Id.* at 850.

16. Sue Anna Moss Cellini, *The Proposed Victims' Rights Amendment to the Constitution of the United States: Opening the Door of the Criminal Justice System to the Victim*, 14 ARIZ. J. INT'L & COMP. L. 839, 847 (1997); see *also* Abramchayev, *supra* note 14, at 849–855; Aynes, *supra* note 5, at 69–73.

17. Abramchayev, *supra* note 14, at 851.

18. *See id.* at 852.

19. *Id.*

20. *See id.* at 854–55.

suggestion that the social contract foundations of our American government should be interpreted as requiring those in power to provide the citizenry with protection from the harms of others. Whether analyzed under constitutional or tort-based principles, the claim that citizens possess an enforceable right to government protection rests on very shaky ground.

B. Constitutional Constraints on a Right to Protection

The Supreme Court has made clear that the social contract theory does not impose a constitutional duty on the government to protect its citizens from the private harm of others. First, in *DeShaney v. Winnebago County Department of Social Services*, the Supreme Court examined whether a county child protective services department could be constitutionally liable for failing to take action to protect a young child, Joshua DeShaney.[21] Joshua was repeatedly beaten by his father, Randy DeShaney.[22] Over a two-year period, county workers were aware of and even responded to reports that Randy was physically abusing Joshua, and at one point temporarily removed Joshua from his father's custody.[23] However, when Joshua was returned to his father's care, case workers did not take any further action when evidence indicated that Randy continued to abuse Joshua.[24] Randy eventually beat his four-year-old son so severely "that he fell into a life-threatening coma[,]"[25] suffered permanent brain damage, and was confined to a medical institution for the rest of his life.[26]

Joshua's mother brought an action against the county and its child services employees, alleging that they "had deprived Joshua of his liberty without due process of law, in violation of his rights under the Fourteenth Amendment, by failing to intervene to protect him against a risk of violence at his father's hands of which they knew or should have known."[27] The Supreme Court rejected her argument, holding that the Fourteenth Amendment, which provides that the "State [shall not] deprive any person of life, liberty, or property, without due process of law[,]"[28] did not impose on the states a duty to protect an individual from the harms caused by private actors.[29] Impliedly rejecting any suggestion that under a social contract theory the government has a specific duty to protect citizens from

21. DeShaney v. Winnebago Cnty. Dept. of Soc. Serv's, 489 U.S. 189 (1989).
22. *Id.* at 191.
23. *Id.* at 192.
24. *Id.* at 192–93.
25. *Id.* at 193.
26. *Id.*
27. *Id.*
28. U.S. CONST. amend. XIV, § 1.
29. *DeShaney*, 489 U.S. at 195.

"invasion[s] by private actors,"[30] the Court noted that the Due Process Clause

> is phrased as a limitation on the State's power to act, not as a guarantee of certain minimal levels of safety and security. It forbids the State itself to deprive individuals of life, liberty or property without "due process of law," but its language cannot fairly be extended to impose an affirmative obligation on the State to ensure that those interests do not come to harm through other means.[31]

The Fourteenth Amendment's Due Process Clause, in concert with the Fifth Amendment's Due Process Clause, did not confer an "affirmative right to governmental aid, even where such aid may be necessary to secure life, liberty, or property interests of which the government itself may not deprive the individual."[32] Therefore, despite the fact that Joshua received care from

---

30. *Id.*

31. *Id.*

32. *Id.* at 196 (citations omitted). The Court did acknowledge that in a narrow set of circumstances, a "duty [to protect] may arise out of certain 'special relationships' created or assumed by the State with respect to particular individuals." *Id.* at 197. Under this "special relationship" exception, "when the State takes a person into custody and holds him there against his will, the Constitution imposes on it a corresponding duty to assume some responsibility for his safety and general well-being." *Id.* at 199–200. However, this duty of protection is not based on "the State's knowledge of the individual's predicament or from [the State's] expressions of intent to help him, but from the limitation which [the State] has imposed on his freedom to act on his own behalf." *Id.* at 200. Hence, most special relationship cases arise where an individual is in some form of government custody. *Id.*

It is often difficult for plaintiffs to successfully assert a special relationship claim. For example, in *Jones v. Phyfer*, a rape victim brought a federal civil rights action against state agency employees, "alleging that their failure to warn her about, and protect her from injury caused by[] the released [defendant] violated her constitutional rights." 761 F.2d 642, 642 (11th Cir. 1985). The victim claimed that because she had initially been victimized by the defendant, and was the reporting witness in the case against him, it should have been foreseeable to the state that the defendant would again attempt to harm her. *Id.* at 646. The court rejected her arguments, contending that they were insufficient to "impose a duty on the state to protect the plaintiff," or to even warn her of the defendant's release. *Id.* at 647; *see also* Gatlin v. Green, 227 F. Supp. 2d 1064, 1076 (D. Minn. 2002) (declining to extend special relationship exception to relationship between informant and police; rather, exception was meant to be limited to prison, "prison-like," and other custodial situations). *But see* Garrett v. Belmont Cnty. Sheriff's Office, 374 F. App'x 612, 618 (6th Cir. 2010) (special relationship exception permitted where woman committed suicide after being released from a county mental institution where county knew of woman's suicidal tendencies); Ex rel Johnson v. South Carolina Dept. of Soc. Serv's, 597 F.3d 163 (4th Cir. 2010) (special relationship exception applicable to situation in which foster child suffered sexual abuse after being taken into state custody).

Along with the "special relationship" exception, a number of courts have recognized that where the state has exacerbated or created the danger which results in harm to a citizen a Fourteenth Amendment substantive due process claim may stand. *See, e.g.*,

state actors who had knowledge of his father's abusive tendencies, there were no constitutional grounds upon which Joshua's mother could hold those public employees liable for the severe harm Joshua suffered.[33]

In *Town of Castle Rock v. Gonzales*, the Court reinforced the *DeShaney* holding and further narrowed the scope of protection permitted under the Fourteenth Amendment's Due Process Clause.[34] In *Gonzales*, Jessica Gonzales was in the process of divorcing her husband, Simon.[35] She had obtained a restraining order against him that limited his ability to come to their former marital home and specifically delineated when he could spend time with their three daughters.[36] The restraining order also contained mandatory language that indicated that the police were required to arrest violators of the order.[37] When Jessica realized her husband had taken their daughters in violation of the order, she contacted the police several times over a period of five hours, asking them to arrest her husband.[38] The police took no direct action in response to Jessica's repeated pleas for help, and instead merely suggested that she wait to see if her husband brought the girls home.[39] Later that evening, Simon showed up at the police station with a semi-automatic weapon he purchased earlier that day, opened fire, and was shot by police in the ensuing cross-fire.[40] Officers found the bodies of his three daughters, whom he had killed earlier that evening, in the cab of his pickup truck.[41] Jessica subsequently brought an action against the city and police for failing to enforce the restraining order.[42]

Okin v. Village of Cornwall-on-Hudson Police Dept., 577 F.3d 415, 438 (2d Cir. 2009); Pena v. Deprisco, 432 F.3d 98, 108 (2d Cir. 2005).

33. *DeShaney*, 489 U.S. at 201.

34. Town of Castle Rock v. Gonzales, 545 U.S. 748 (2005).

35. *Id.* at 751.

36. *Id.*

37. The restraining order issued the following command to law enforcement officials:

> You *shall* use every reasonable means to enforce this restraining order. You *shall* arrest, or, if an arrest would be impracticable under the circumstances, seek a warrant for the arrest of the restrained person when you have information amounting to probable cause that the restrained person has violated or attempted to violate any provision of this order and the restrained person has been properly served with [notice] of this order or has received actual notice of the existence of this order.

Id. at 752 (emphasis added). The language in the restraining order mirrored language that appeared in Colorado statutes mandating that police arrest individuals who violate the terms of a restraining order. *See* COLO. REV. STAT. §§ 18-6-803.5(3)(a), (b).

38. *Gonzales*, 545 U.S. at 753–54.

39. *Id.*

40. *Id.* at 754.

41. *Id.*

42. *Id.*

In contrast to the action in *DeShaney*, which was premised on the substantive due process clause of the Fourteenth Amendment,[43] Jessica framed her action against the city under the Fourteenth Amendment's procedural due process clause.[44] First, she asserted that she possessed a property interest in the enforcement of the restraining order.[45] Second, she claimed that the town's policy of tolerating police unresponsiveness to reports of restraining order violations represented a deprivation of her property without due process.[46] The Supreme Court disagreed. In similar fashion to its *DeShaney* ruling, the Court was unwilling to interpret the Fourteenth Amendment as imposing a constitutional duty of protection on state actors.[47] Not only did the Court refuse to rule that the State of Colorado had created a statutory property interest in the enforcement in the restraining order,[48] but it was equally unwilling to consider that *even if* the state legislature had intended to create such a property right, that the right was sufficient to rise to a level warranting Constitutional protection.[49] Instead, the Court reasoned that any existing mandatory enforcement duty benefitted society as a whole, rather than the individual holding the restraining order.[50] In like fashion to Joshua DeShaney and his mother,

---

43. DeShaney v. Winnebago Cnty. Dep't of Soc. Serv's, 489 U.S. 189, 195 (1989).

44. In *DeShaney*, and as noted by the Court in *Gonzales*, Joshua and his mother had raised the issue of whether the state's child protection statutes had created an entitlement in Joshua that could not be deprived without procedural due process of law. 489 U.S. at 195 n.2; *see also Gonzales*, 545 U.S. at 755. However, because this issue was raised in an untimely manner, the *DeShaney* Court declined to address it. 489 U.S. at 195 n.2. Hence, the question of whether the procedural due process clause might permit an individual to sue a state actor for failing to provide protective services remained an open question for review in *Gonzales*. 545 U.S. at 755.

45. *Gonzales*, 545 U.S. at 755.

46. *Id.*

47. *Id.* at 768–69.

48. *Id.* at 758–66. In *Board of Regents of State Colleges v. Roth*, the Supreme Court ruled that states could create property interests protected under the procedural due process clause of the Constitution. 408 U.S. 564, 577 (1972). However, the *Gonzales* Court was unwilling to rule that the mandatory enforcement language which appeared in Colorado law and on Jessica Gonzales' restraining order were sufficient to create a *Roth*-type of constitutionally protected property interest. *Gonzales*, 545 U.S. at 758–62.

49. *Id.* at 766–68.

50. Justice Scalia noted that

> Making the actions of government employees obligatory can serve various legitimate ends other than the conferral of a benefit on a specific class of people. The serving of public rather than private ends is the normal course of the criminal law because criminal acts, "besides the injury [they do] to individuals, . . . strike at the very being of society. . . ." 4 W. Blackstone, COMMENTARIES ON THE LAWS OF ENGLAND 5 (1769).

Id. at 765.

Jessica Gonzales was left without a constitutional remedy against the Town of Castle Rock for her daughters' deaths.[51]

The *DeShaney* and *Gonzales* cases teach that government protection from the wrongful acts of others does not fall within the spectrum of constitutionally protected individual rights.[52] This is not to say, however, that the individual states or Congress are barred from imposing a statutory tort duty of protection on government actors. In both *DeShaney* and *Gonzales,* the Court begrudgingly suggested as much. In noting that the Fifth and the Fourteenth Amendments were not designed to require the States to protect citizens from one another, the *DeShaney* Court noted that "[t]he Framers were content to leave the extent of governmental obligation in [this] area to the democratic political processes."[53] Similarly, in *Gonzales,* the Court stated that the States are not "powerless to provide victims with personally enforceable remedies."[54] Therefore, there is certainly room within the law for governments to craft a protection right for victims. However, even where the states and the federal government have created tort based or legislatively enacted protection rights for citizens, an individual's ability to enforce this right or seek redress for its violation is precarious and inconsistent.[55]

---

51. *See id.* at 768–69; DeShaney v. Winnebago Cnty. Dep't of Soc. Serv's, 489 U.S. 189, 195–97 (1989).

52. In concert with the special relationship and state-created danger exceptions, *see supra* note 32, one additional narrow exception exists. Included within the Fourteenth Amendment is the prohibition that States shall not "deny to any person within its jurisdiction the equal protection of the laws." U.S. CONST., amend. XIV, § 1. In *DeShaney,* the Court aptly noted that "[t]he State may not, . . . selectively deny its protective services to certain disfavored minorities without violating the Equal Protection Clause." 489 U.S. at 197 n.3 (citations omitted); *see also* Moore v. City of Chicago Heights, No. 063452, 2010 WL 148623 (N.D. Ill. Jan. 12, 2010) (equal protection claim that city had a policy, pattern and practice of treating victims of domestic violence with less priority than victims of other crimes, survived motion to dismiss); Elliot-Park v. Manglona, 592 F.3d 1003 (9th Cir. 2010) (police officer's failure to investigate and arrest drunk driver due to racial preferences toward the driver and against the injured party gave rise to an equal protection claim); Price-Cornelison v. Brooks, 524 F.3d 1103 (10th Cir. 2008) (police denied qualified immunity in equal protection action brought by lesbian victim of domestic violence).

53. *DeShaney,* 489 U.S. at 196; *see also id.* at 202 ("A State may, through its courts and legislatures, impose . . . affirmative duties of care and protection upon its agents as it wishes.").

54. *Gonzales,* 545 U.S. at 768. The Court went on to cite a variety of cases in which state courts had indicated that cities and police officers could be held liable for failing to provide certain protection based services. *Id.* at 769 n.15; *see also* Linda R. S. v. Richard D., 410 U.S. 614, 617 n.3 (1973) (citations omitted) (noting that legislative bodies "may enact statutes creating legal rights, the invasion of which creates standing even though no injury would exist without the statute").

55. *See* Amy Felman, *The Special Duty Doctrine: A Just Compromise,* 31 ST. LOUIS U. L.J. 409, 420–25 (1987).

C. State Tort Limits on the Victim's Protection Right

There are certainly reported cases in which state courts have acknowledged that state actors can and should be held liable for harm suffered by citizens at the hands of another.[56] However, there is no ironclad formula to ensure that liability will be imposed when a victim's right to protection is violated. The sovereign immunity,[57] public duty,[58] and special duty doctrines[59] bog down a victim's right to claim protection, as do other statutory limits on state actor liability. The end result is that even where a legitimate claim for protection might exist, there is little guarantee that the claim can clear the common law and statutory hurdles limiting its enforcement.

1. Sovereign Immunity, the Public Duty Doctrine, and the Special Duty Doctrine

Sovereign immunity and the public duty doctrine have traditionally limited state actor accountability. Hailing from our English common law roots, the sovereign immunity doctrine originally served as a means of protecting all state actors from any form of liability. As rationalized by Justice Holmes, "[a] sovereign is exempt from suit, not because of any

56. *See* Peschel v. City of Missoula, 664 F. Supp. 2d 1149 (D. Mont. 2009) (where individual was in custody of police and was subsequently denied medical care, special relationship exception claim can survive motion for summary judgment); Kaho'Ohanohano v. Dep't of Human Serv's, State of Hawaii, 178 P.3d 538 (Haw. 2008) (Department of Human Services owed a duty of care to minor once abuse of child was confirmed by investigators); Pile v. Brandenburg, 215 S.W.3d 36 (Ky. 2007) (police officer liable for negligent performance of ministerial act, where woman was killed in collision because officer failed to prevent a handcuffed and intoxicated prisoner left alone in vehicle from taking control of car); Radke v. County of Freeborn, 694 N.W.2d 788 (Minn. 2005) (state Department of Human Services could be charged with negligence in investigation of child abuse reports); Cockerham-Ellerbee v. Town of Jonesville, 626 S.E.2d 685 (N.C. Ct. App. 2006) (officer's specific promise of protection to plaintiff and her daughter was sufficient to permit plaintiff to sue city for officer's failure to fulfill promise); Estate of Graves v. City of Circleville, 922 N.E.2d 201 (Ohio 2010) (wanton and reckless actions by city employees can give rise to claims of negligence).

57. *See Estate of Graves*, 922 N.E.2d 201 (stating, in passing, that sovereign immunity is the "prov[ision of] broad statutory immunity to political subdivisions and their employees, subject to certain exceptions.") (citations omitted).

58. *See Cockerham-Ellerbee*, 626 S.E.2d at 687–88 ("Generally, the public duty doctrine bars negligence claims by individuals against a municipality or its agents acting in a law enforcement role for failure to provide protection to that person from the criminal acts of third party.") (citations omitted).

59. *See Carcraft*, 279 N.W.2d at 806 ("'Special duty' is nothing more than convenient terminology, . . . for the ancient doctrine that once a duty to act for the protection of others is voluntarily assumed, due care must be exercised even though there was no duty to act in the first instance.") (citations omitted); *accord Radke*, 694 N.W.2d at 793.

formal conception or obsolete theory, but on the logical and practical grounds that there can be no legal right as against the authority that makes the law on which the right depends."[60] Most states, as well as the federal government, however, have departed from a rigid adherence to sovereign immunity,[61] and have exposed state employees to liability for their performance of select tasks and duties. The softening of sovereign immunity is most often exhibited in states' invocation of the public duty doctrine.[62]

The public duty doctrine generally provides that a state actor will not be liable for obligations owed to the general public, such as police protection or fire prevention services.[63] Therefore, while the public duty doctrine

---

60. Kawananakoa v. Polyblank, 205 U.S. 349, 353 (1907).

61. *See* Hicks v. State, 544 P.2d 1153, 1155–57 (N.M. 1976), *superseded by statute*; Mayle v. Penn. Dep't of Highways, 388 A.2d 709, 710 (Pa. 1978); Jeffrey D. Hickman, Note, *It's Time to Call 911 For Government Immunity*, 43 CASE W. RES. L. REV. 1067, 1074–75 (1993); L. Stephens & Bryan P. Harnetiaux, *The Value of Government Tort Liability: Washington State's Journey From Immunity to Accountability*, 30 SEATTLE U. L. REV. 35, 36 (2006); Kelly Mahon Tullier, *Governmental Liability for Negligent Failure to Detain Drunk Drivers*, 77 CORNELL L. REV. 873, 878 (1992); Mark L. Van Valkenburgh, *Massachusetts General Laws Chapter 258, § 10: Slouching Toward Sovereign Immunity*, 29 NEW ENG. L. REV. 1079, 1081–83 (1995). The Federal Tort Claims Act also imposes liability on the government for select tortuous activities of employees. *See* 28 U.S.C. §§ 1364, 1402, 2401–2402, 2411–2412, 2671–2680 (2006). States have also passed similar acts, whereby citizens are permitted to raise negligence claims against state and municipal employees but only for a narrowly prescribed set of cases. *See, e.g.*, ALASKA STAT. §§ 09.50.250 to 09.50.300 (2008); COLO. REV. STAT. §§ 24-10-101, 24-10-106(1.5)(a-c) to 24-10-106(2-4) (2010); FLA. STAT. § 768.28 (2005 & Supp. 2010); GA. CODE ANN. §§ 50-21-21 to 50-21-37 (2009); HAW. REV. STAT. §§ 662-1 to 662-19 (1993 & Supp. 2007); IDAHO CODE ANN. §§ 6-901 to 6-929 (2010); IND. CODE ANN. §§ 34-13-3-1 to 34-13-3-25 (2008); IOWA CODE §§ 669.1 to 669.24 (2000 & Supp. 2010); KAN. STAT. ANN. §§ 75-6101 to 75-6120 (1997); KY. REV. STAT. ANN. §§ 44.070 to 44.073 (2007); MD. CODE ANN. CTS. & JUD. PROC. §§ 5-502 to 5-524 (2009); MD. CODE ANN. ST. GOV'T §§ 12-101 to 12-110 (2009); MASS. GEN. LAWS ch. 258, §§ 1–14 (2001 & Supp. 2010); MINN. STAT. § 3.736 (2002); MISS. CODE ANN. §§ 11-46-1 to 11-46-23 (2002 & Supp. 2009); MONT. CODE ANN. §§ 2-9-101 to 2-9-112 (2009); NEV. REV. STAT. §§ 41.0305 to 41.0344 (2009); N.H. REV. STAT. ANN. §§ 541-B:1 to 541-B:21 (2007 & Supp. 2009); N.M. STAT. §§ 41-4-1 to 41-4-13 (2010); N.D. CENT. CODE. §§ 32-12.1-01 to 32-12.1-04 (2010); OHIO REV. CODE ANN. § 2743.02(A)(3)(a) (2008); OKLA. STAT. tit. 51 §§ 151–172 (2008); OR. REV. STAT. §§ 30.260-30.302 (2009); R.I. GEN. LAWS §§ 9-31-1 to 9-31-13 (1997 & Supp. 2009); S.C. CODE ANN. §§ 15-78-10 to 15-78-220 (2003 & Supp. 2009); S.D. CODIFIED LAWS §§ 3-21-1 to 3-21-12 (2004); TEX. CIV. PRAC. & REM. CODE ANN. §§ 101.001–101.107 (2005 & Supp. 2010); VT. STAT. ANN. tit. 12 §§ 5601–5606 (2002 & Supp. 2009); VA. CODE ANN. §§8.01-195.1 to 8.01-195.9 (2009 & Supp. 2010).

62. *See* Stephens & Harnetiaux, *supra* note 61, at 35–36.

63. The public duty doctrine was first articulated by the Supreme Court in *South v. Maryland*, 59 U.S. 396 (1855). The Court indicated that while the police may have had a general duty to protect the public, that broad duty did not translate into liability for failing to

wedged open the tightly closed door of sovereign immunity, it still continues to severely limit the types of claims that can be brought against public officers for the dereliction of their duties.[64] The Supreme Court's reasoning in both *DeShaney* and *Gonzales* echoes strains of the public duty doctrine. In neither case was the Court willing to extend from some broad general-protection duty owed to the public a more specific duty owed to a particular citizen. Whether it was a county's creation of a department of social services to oversee the well being of children,[65] or a legislature's command that police officers "shall" arrest violators of restraining orders,[66] the Court made clear that these orders were made for the benefit of the community as a whole and not for the individual benefit of Joshua DeShaney or Jessica Gonzalez and her daughters.[67]

The narrowness of the public duty doctrine has been criticized by academics[68] and softened by courts and legislatures with the special duty doctrine.[69] The special duty doctrine holds that even within the strict confines of the public duty doctrine, a state actor can be liable for private harm where a special duty was established between that specific state actor and the individual requesting assistance.[70] On its face, the special duty

---

protect a specific individual. *Id.* at 403; *see also* Shepard v. Bradford, 721 So. 2d 1049 (La. Ct. App. 1998) (under public duty doctrine city not liable for failing to protect plaintiff attacked on public basketball court by a defendant with a history or propensity for violence); Cynthia Zeliner Mackinnon, *Negligence of Municipal Employees: Redefining the Scope of Police Liability*, 35 U. FLA. L. REV. 720, 725 (1983); Miami-Dade County v. Miller, 19 So. 3d 1037 (Fla. Dist. Ct. App. 2009) (public duty doctrine barred suit against city for failing to assign police officers to patrol bus stop where victim was subsequently assaulted); Partain v. Oconee County, 667 S.E.2d 132 (Ga. Ct. App. 2008) (public duty doctrine barred claim against deputy who failed to arrest a drunk driver who subsequently killed another driver); Raas v. State, 729 N.W.2d 444 (Iowa 2007) (public duty doctrine barred claim of fisherman who was attacked by two escaped prisoners).

64. *See* Myers v. McGrady, 628 S.E.2d 761, 766 (N.C. 2006) ("The public duty doctrine is a separate rule of common law negligence that may limit tort liability, even when the State has waived sovereign immunity.").

65. DeShaney v. Winnebago County Dep't of Social Servs., 489 U.S. 189, 195 (1989).

66. Town of Castle Rock v. Gonzales, 545 U.S. 748 (2005).

67. *See id.* at 765.

68. *See* Aaron R. Baker, *Untangling the Public Duty Doctrine*, 10 ROGER WILLIAMS U. L. REV. 731 (2005); David Basil, *A Primer on the Public Duty Doctrine as Applied to Police Protection*, 37 URB. LAW. 403, 405 (2005); Suzanne M. Dardis, Gleason v. Peters: *An Application of the Public Duty Doctrine as a Judicial Resurrection of Sovereign Immunity*, 43 S.D. L. REV. 706 (1998); Harvard Law Review Assoc., *Police Liability for Negligent Failure to Prevent Crime*, 94 HARV. L. REV. 821 (1981); Hickman, *supra* note 61, at 1067; G. Kristian Miccio, *Exiled from the Province of Care: Domestic Violence, Duty and Conceptions of State Accountability*, 37 RUTGERS L.J. 111, 124–25 (2005); G. Kristian Miccio, *If Not Now, When? Individual and Collective Responsibility for Male Intimate Violence*, 15 WASH. & LEE. J. CIVIL RTS. & SOC. JUST. 405, 430–38 (2009).

69. *See* Cracraft v. City of St. Louis, 279 N.W.2d 801, 806 (1979).

70. *See* Leone v. City of Chicago, 619 N.E.2d 119 (Ill. 1993) (where plaintiff can

doctrine could further the argument that a victim's statutory right to protection from the accused should be enforced against state actors.[71] If a statute states "the victim has a right to be reasonably protected from the accused,"[72] then one could argue that a specific duty has been imposed upon law enforcement to protect a particular class of individuals, the victims of crimes, and therefore the police should be liable for the failure to provide that protection. However, in practice, the special duty doctrine tends to be applied narrowly[73] and with varied results.[74]

---

prove four elements of special duty exception police and city liable for injuries sustained by plaintiff struck by car at traffic stop); Cracraft, 279 N.W.2d at 806–07 (identifying four factors to be considered in determining when a special duty arises); Cockerham-Ellerbee v. Town of Jonesville, 626 S.E.2d 685 (N.C. Ct. App. 2006) (special duty exception satisfied where plaintiff had protective order and was given specific assurances of protection by police.); Ezell v. Cockerell, 902 S.W.2d 394, 402 (Tenn. 1995) (identifying three factors to determine whether a special duty of care exists); *see also* David A. Aaby, *The Scope of the Public Duty/Special Duty Doctrine in Illinois: Municipal Liability for Failure to Provide Police Protection*, 10 N. ILL. U. L. REV. 269, 288–91 (1990); Basil, *supra* note 68, at 406–07; Felman, *supra* note 55, at 420–21; Gerald P. Krause, Comment, *Municipal Liability: The Failure to Provide Adequate Police Protection-The Special Duty Doctrine Should be Discarded*, 1984 WIS. L. REV. 499, 508–10 (1984); Courtney E. Nutall, *Torts-Matthews v. Pickett County: The Public Duty Doctrine and its Special Duty Exception in the Face of the Governmental Tort Liabilities Act*, 30 U. MEM. L. REV. 457, 466–67 (2000); Alexander B. Punger, *Protecting the Greater Good: A Critique of the Public Duty Doctrine as Applied in* Murray v. County of Pearson, 88 N. C. L. REV. 694, 699 (2010); Tullier, *supra* note 61, at 889–93.

71. *See generally* Aaby, *supra* note 70, at 288–91.

72. *See, e.g.*, ALASKA CONST. art. I, § 24; ARIZ. CONST. art. II, § 2.1(A)(1); CAL. CONST. art. I, § 28(a); CONN. CONST. art. I, § 8(b)(3); ILL. CONST. art. I, § 8.1(a)(7); MICH. CONST. art. I, § 24(1); MO. CONST. art. I, § 32(6); N.M. CONST. art. II, § 24(3); OHIO CONST. art. I, § 10a; OR. CONST. art. I, § 43(1)(a); S.C. CONST. art. I, § 24(A)(6); TENN. CONST. art. I, § 35(2); TEX. CONST. art. I, § 30(2); VA. CONST. art. I, § 8-A(1); WIS. CONST. art. I, § 9m; 18 U.S.C. § 3771(a)(1); ALASKA STAT. § 12.61.010(4) (2008); COLO. REV. STAT. § 24-4.1-302.5(a) (2010); HAW. REV. STAT. § 801D-4(3) (Supp. 2007); IND. CODE. ANN. § 35-40-5-1(2) (Supp. 2010); KAN. STAT. ANN. § 74.7333(7) (1992); MD. CODE ANN. CRIM. PROC., § 11-1002(4) (LexisNexis 2008); MASS. GEN. LAWS ch. 258B § 3(d) (2001 & Supp. 2010); MISS. CODE. ANN. § 99-36-5(1)(a) (2007); NEB. REV. STAT. § 81-1848(c) (2008); N.H. REV. STAT. ANN. § 21-M:8-k(II)(c) (2008 & Supp. 2009); N.Y. EXEC. LAW § 646-a (2005 & Supp. 2010); OKLA. STAT. tit. 19, § 215.33(A)(2) (2000 & 2010); R.I. GEN. LAWS § 12-28-3(3) (2002 & Supp. 2009); S.D. CODIFIED LAWS § 23A-28C-1(4) (2004 & Supp. 2010); TEX. CODE CRIM. PROC. ANN. § 56.02(a)(1) (2006 & Supp. 2010); WASH. REV. CODE. § 7.69.030(4) (2008 & Supp. 2009); WYO. STAT. ANN. § 14-6-504(c) (2009).

73. *See* Punger, *supra* note 70, at 699.

74. *See* Felman, *supra* note 55, at 420–25.

2. A State Case Study: Massachusetts, Promises of Protection, and Tort Limits

The Commonwealth of Massachusetts provides but one example of how the special duty doctrine cannot guarantee a protection right to crime victims. Where states have embodied the special duty doctrine in their state tort claim acts, they often do so by carving out only the most explicit of exceptions to the public duty doctrine. Massachusetts fits this model in that its state tort claims act starts out broadly, imposing liability on public employers for the negligent acts of their employees "in the same manner and to the same extent as a private individual under like circumstances."[75] The statute then narrows the scope of liability by detailing several activities for which public actors will not be liable.[76] The Massachusetts courts have identified this legislative reduction of liability as an embodiment of the public duty doctrine.[77]

Three of the public duty doctrine exceptions are of particular relevance to Massachusetts's crime victims and their ability to enforce a right to protection. These three exceptions detail that public actors will not be liable for (1) "failure to provide adequate police protection," (2) "any claim based upon the release, parole, furlough or escape of any person . . . from the custody of a public employee or employer or their agents, unless gross negligence is shown in allowing such release, parole, furlough or escape" or (3) "any claim based on an act or failure to act to prevent or diminish the harmful consequences of a condition or situation, including . . . violent or tortious conduct of a third person . . . not originally caused by the public employer or any other person acting on behalf of the public employer."[78] Therefore, the Massachusetts statute presents a bit of an open-the-door then close-the-door approach to public actor liability. The state tort claims act begins with an all-inclusive approach, but then narrows the circumstances under which public employees can be sued.[79]

Within the broad swath of immunity granted to Massachusetts's state actors, the state's tort claims act statute nonetheless includes two important counter-exceptions that exemplify the special duty doctrine. First, state employees can be liable for "explicit and specific assurances of safety or

---

75. Mass. Gen. Laws ch. 258 § 2 (2001 & Supp. 2010).

76. *See id.* at § 10.

77. In *Jean W. v. Commonwealth*, the Massachusetts Supreme Court announced its intention to abolish the common law public duty rule, but invited the Legislature "to respond to this anticipated change by passing additional limitations on liability." 610 N.E.2d 305, 307 (Mass. 1993). The legislature did so by passing section 10 of the state tort claims act. The Massachusetts courts subsequently identified this statutory change as enshrining the public duty doctrine in statute. *See* Ford v. Grafton, 693 N.E.2d 1047, 1052–53 (Mass. 1998); Lawrence v. City of Cambridge, 664 N.E.2d 1, 2–3 (Mass. 1996).

78. MASS. GEN. LAWS ch. 258, §§ 10(h)–(j).

79. *See id.* at §§ 2, 10.

assistance . . . made to a direct victim or member of his family . . . provided the [subsequent] injury resulted in part from [the victim's] reliance on those assurances."[80] Second, liability will be imposed where "the intervention of a public employee causes injury to the victim or places the victim in a worse position than he was in before the intervention."[81] Massachusetts law also specifically grants crime victims the right "to receive protection from the local law enforcement agencies from harm and threats of harm arising out of their cooperation with law enforcement and prosecution efforts."[82] Taken in concert with Massachusetts's special duty exceptions, one might conclude that should a victim suffer harm from the offender during the course of the state's investigation and prosecution of a crime, state actors could be held liable. Unfortunately, the commonwealth's highly fact-sensitive application of the special duty doctrine, coupled with other statutory limits in Massachusetts law, fail to guarantee victims protection or relief.[83]

A victim's success in asserting a special duty claim is highly predicated on the facts of the particular case before the court.[84] For example, in interpreting the special duty language incorporated into the Massachusetts statute, courts have focused on the requirement that the state actor make "explicit and specific assurances of safety or assistance" to the victim.[85] So doing, the Massachusetts Supreme Court has held that "by 'explicit' the Legislature meant a spoken or written assurance, not . . . [assurances] implied from the conduct of the parties or the situation, and by 'specific' the terms of the assurance must be definite, fixed and free from ambiguity."[86] A victim's success in claiming she received "explicit and

---

80. *Id.* at § 10(j)(1).

81. *Id.* at § 10(j)(2). These two exceptions to the immunity granted to state officials closely mirror the special relationship and state created danger doctrines often invoked in Fourteenth Amendment substantive due process claims. *See supra* note 32.

82. MASS. GEN. LAWS ch. 258B, § 3(d).

83. *See id.* at §§ 2, 10.

84. *See generally* Basil, *supra* note 68, at 407. As Mr. Basil notes, each state has come up with its own formulation of what is required to successfully claim that a special duty has arisen between a state employee and a member of the public. *Id.* at 407; *see, e.g.*, Doe v. Calumet City, 641 N.E.2d 498, 504 (Ill. 1994) (articulating a four part test); Serviss v. State, Dep't of Natural Resources, 711 N.E.2d 95, 99 (Ind. Ct. App. 1999) (articulating a three part test) *vacated on other grounds*, 721 N.E.2d 234 (Ind. 1999); Cracraft v. City of St. Louis Park, 279 N.W.2d 801 (Minn. 1979) (identifying four factors to be considered in determining when a special duty arises); Summers v. Harris Construction, 381 S.E.2d 493, 496 (S.C. Ct. App. 1989) (articulating a six part test); Ezell v. Cockerell, 902 S.W.2d 394, 402 (Tenn. 1995) (identifying three factors to determine whether a special duty of care exists).

85. MASS. GEN. LAWS ch. 258, § 10(j)(1).

86. Lawrence v. City of Cambridge, 664 N.E.2d 1, 3 (Mass. 1996); *see also* Ariel v. Town of Kingston, 867 N.E.2d 367, 370 (Mass. Ct. App. 2007) (citations omitted); Ford v. Town of Grafton, 693 N.E.2d 1047, 1054 (Mass. 1998) (citations omitted).

specific assurances of safety" from state employees will therefore always be very fact specific.[87]

For example, in *Lawrence v. City of Cambridge*, the Supreme Court of Massachusetts found that the plaintiff-victim had sufficiently pled that police officers and the city had made him "explicit and specific assurances of safety."[88] The plaintiff had been robbed at gunpoint one evening as he was closing his store.[89] He agreed to testify against his assailants at a grand jury hearing and the police promised they would protect him when he closed his store in the evenings.[90] Relying on this promise, the plaintiff felt safe to return to work.[91] An officer was present at the plaintiff's place of work for several days, but the evening before the plaintiff was meant to testify at the grand jury hearing, the police were absent.[92] The police had not informed the plaintiff that they were going to stop protecting him or that they were unable to protect him on that specific evening.[93] As the plaintiff was leaving work, someone shot him in the face.[94] Based on these facts, the court determined that the plaintiff had not only shown that the police department's promise was explicit, as the police had verbally promised protection to the plaintiff, but that their promise was also specific.[95] According to the court, the officers' promise that they would "provide the plaintiff protection 'when [he] closed the store at night' . . . specifie[d] some of the most important terms of the assurance—when and where."[96] Hence, *Lawrence* suggested that the special duty language appearing in the state tort claims act could be read with a measure of breadth.[97] Subsequent Massachusetts cases, however, have not construed the special duty exception so broadly.

In *Ford v. Town of Grafton*, Catherine Ford received repeated death threats for nearly two years from her ex-husband, James Davison, despite holding a restraining order against him.[98] Catherine repeatedly asked the police to arrest her ex-husband when he threatened her safety, the safety of her family members, and further violated the terms of the restraining order by harassing her.[99] However, the police consistently declined to take any action[100] despite state law that mandated arrest when police had probable

87. *See* Basil, *supra* note 68, at 407.
88. *Lawrence*, 644 N.E.2d at 3.
89. *Id.* at 2.
90. *Id.*
91. *Id.*
92. *Id.*
93. *Id.*
94. *Id.*
95. *Id.* at 4.
96. *Id.* (alterations in original).
97. *Id.*
98. Ford v. Town of Grafton, 693 N.E.2d 1047, 1049–51 (Mass. Ct. App. 1998).
99. *Id.*
100. *Id.*

cause that a restraining order was being violated.[101] Instead, the police erroneously asserted that they could not arrest James unless he caused Catherine actual physical harm or the police specifically saw him violating the restraining order.[102]

James's threats and violence against Catherine eventually escalated to such a level that he shot her, rendering her a quadriplegic.[103] Catherine subsequently brought an action against the town and its police force.[104] She grounded her claim on the portion of the state's tort claim act that imposed liability on public employees for "any claim based upon explicit and specific assurances of safety or assistance"[105] and state law which mandated that police arrest violators of restraining orders.[106] The court rejected her claim, contending that to the extent that Catherine received any explicit or specific assurances from the police, none were for protection.[107] Rather, "the town assured [Catherine] Ford and her family it could not take any action until it saw [James] Davidson violating the protective order or he actually caused [her] harm."[108] That the police were wrong on this issue was of no matter.[109] Likewise, the court rejected any argument that the mandatory arrest language in the restraining order statute created any explicit or specific promise of protection. While the court acknowledged that officers are mandated to "act in accordance with the statute[,]" the court found "no language in the statute that holds officers liable for failing

---

101. *Id.* at 1049–52, 1054–55. In particular, the Massachusetts statute guided that police use all reasonable means to prevent domestic abuse, including

> arresting any person whom the officer has probable cause to believe has committed a felony; . . . arresting any person who has committed, in the officer's presence, a misdemeanor which involves abuse; . . . arresting any person whom the officer has probable cause to believe has committed a misdemeanor [related to domestic violence].

Id. at 1052 n.4 (citation omitted). Despite this statutory language, the police informed Caroline that there was nothing they could do about James's harassment, that "they could not babysit her twenty-four hours a day[,]" and that her best option was to buy a gun "because the only way to deal with violence was violence." *Id.* at 1049.

102. *Ford*, 693 N.E.2d at 1050.

103. *Id.* at 1051. After shooting Caroline, James fatally shot himself. *Id.*

104. *Id.* at 1051–52.

105. *Id.* at 1053 (citation omitted).

106. *See id.* at 1052 n.4 (citation omitted); *supra* note 101 (detailing terms of restraining order statute (detailing the terms of the Massachusetts restraining order statute MASS. GEN. LAWS ch. 209A, § 6(4-6) (2007)).

107. *Ford*, 693 N.E.2d at 1054.

108. *Id.*

109. *Id.* at 1050.

to do so."[110] Therefore, Catherine's reliance on the promise of protection in the restraining order was misplaced and could not afford her legal relief.[111]

In *Ariel v. Town of Kingston*, the Massachusetts Appeals Court was similarly unwilling to give a broad construction to the statute's "explicit and specific assurances of safety or assistance" language.[112] There, a driver came upon a road accident where the police were present and directing

---

110. *Id.* at 1054–55.

111. *Id.* Not every state is so parsimonious when examining whether a restraining order imposes a tort-based duty on public employees to provide protection to the holder of the order. For example, in Tennessee, the special duty doctrine is applicable, in part, when "a public official affirmatively undertakes to protect the plaintiff and the plaintiff relies on the undertaking. . . ." *Matthews v. Pickett County*, 996 S.W.2d 162, 165 (Tenn. 1999) (citing *Chase v. City of Memphis*, 971 S.W.2d 380, 385 (Tenn. 1998); *Ezell v. Cockrell*, 902 S.W.2d 394, 402 (Tenn. 1995)). In *Matthews*, Mary Matthews was assaulted, beaten and sexually violated by her estranged husband, Bill Winningham, and subsequently obtained a restraining order against him. 996 S.W.2d at 163. The police nonetheless failed to arrest Bill, despite his repeated violations of the restraining order. *Id.* at 164. Mary's estranged husband eventually burned down her home, and Mary brought an action against the police for failing to enforce the terms of her restraining order. *Id.* In contrast to the Massachusetts court's analysis in *Ford*, the Tennessee Supreme Court determined that Mary Mathew's claim fit within the state's special relationship exception. The court noted:

> [t]he order of protection in this case was not issued for the public's protection in general. The order of protection specifically identified Ms. Matthews and was issued solely for the purpose of protecting her. . . . Ms. Matthews apparently relied on the court's order of protection. She contacted the sheriff's department and requested that it provide her with protection pursuant to the order of protection. Accordingly, the special duty exception to the public duty doctrine is applicable to this case.

Matthews, 996 S.W.2d at 165 (citation omitted).

The state of Tennessee nonetheless imposes other obstacles to a victim's attempt to enforce a protection right. Like Massachusetts, Tennessee grants victims the right to "protection and support with prompt action in the case of intimidation and retaliation from the defendant and the defendant's agents or friends." TENN. CODE ANN. § 40-38-102(a)(2) (2006). However, the state's victims' rights statute undermines the enforceability of this protection right by mandating that a state actor's

> [f]ailure to comply with any provision of [the state's victims' rights laws] shall not create a cause of action or a claim for damages against the state, a political subdivision of the state, a government employee or other official or entity, and no such cause of action shall be maintained.

§ 40-38-108. So while a victim might be able to seek some refuge under the state court's application of the special duty doctrine, *see Matthews*, 996 S.W.2d 162, relief pursuant to the state's victims' rights laws is limited. § 40-38-108.

112. Ariel v. Town of Kingston, 867 N.E.2d 367, 369–70 (Mass. App. Ct. 2007).

traffic.[113] The driver continued through the intersection when she had a green traffic signal and was hit by a driver who she claimed had been negligently directed into the intersection by one of the officers.[114] The court conceded that the "police officers' direction of traffic on a public way constitutes a form of providing police protection to the public."[115] Nonetheless, the court was unwilling to conclude that police traffic direction was either explicit or specific enough to fall within the scope of the statutory language imposing liability on state actors for failure to protect an individual from harm.[116]

While a Massachusetts victim might be stymied in the ability to consistently seek relief under the special duty doctrine, a victim might attempt to assert that the promise of protection found in the state's victims' rights statute[117] is sufficient to fall within the narrow window of liability created in the state's tort claims act.[118] Unfortunately, other provisions in Massachusetts's state tort claims act further preclude relief to crime victims.

The state tort claims act directs that liability shall not be imposed for "any claim based upon an act or omission of a public employee when such employee is exercising due care in the execution of *any statute or any regulation of a public employer*."[119] This statutory language indicates liability will be imposed only in those instances when a public employee's acts fall short of the standards of due care. Therefore, the victim's protection right is not absolute. An officer's good faith efforts, which may nonetheless result in injury to the victim, will be shielded. The Massachusetts courts have also indicated that the immunities listed in the state tort claims act operate in the alternative. Even if a victim could make an argument that her protection claim fell within one of the statute's immunity exceptions, the claim will still be barred if any of the other immunities could shield the state actor from liability.[120]

Finally, specific terms within the Massachusetts's victims' rights laws themselves foreclose the victim's ability to raise a claim against a state employee for failing to comply with the protection duties listed in the

113. *Id.* at 369.

114. *Id.*

115. *Id.* at 370.

116. *Id.*

117. MASS. GEN. LAWS ch. 258B, § 3(d) (2004) (stating that victims have the right "to receive protection from the local law enforcement agencies from harm and threats of harm arising out of their cooperation with law enforcement and prosecution efforts").

118. *Id.* at § 10(j)(1) (stating that liability can be imposed for a "claim based on explicit and specific assurances of safety or assistance . . . made to the direct victim or a member of his family or household by a public employee, provided that the injury resulted in part from reliance on those assurances").

119. MASS. GEN. LAWS. ch. 258, § 10(a) (emphasis added).

120. *See* Brum v. Town of Dartmouth, 704 N.E.2d 1147, 1155–56 (Mass. 1999); Ariel v. Town of Kingston, 867 N.E.2d 367, 371 (Mass. App. Ct. 2007).

statute. Despite granting victims a host of "fundamental rights,"[121] the Massachusetts legislature has nonetheless made clear that "[n]othing [within the victims' rights legislation] shall be construed as creating an entitlement or cause of action on behalf of any person against any public employee, public agency, the commonwealth or any agency responsible for enforcement of rights and provision of services set forth"[122] in the state's victims' rights laws.[123] Hence, a victim's expectation that state law guarantees the direct service of protection from the offender is likely to be satisfied in only the narrowest of situations.

One can extract many lessons from the Commonwealth of Massachusetts. First, even where a victim might be able to raise a special duty claim, success is not guaranteed. The victim will have to prove that the specific facts of her case satisfy the particular manner by which the courts have interpreted the scope of the special duty doctrine. Second, even if state law provides some initial ground upon which a protection right would exist, other statutory provisions often foreclose a victim's ability to enforce that right or seek relief when the right is violated. Hence, most governmental promises of protection should be described as aspirational goals: "We hope, wish, desire, to provide victims with protection, but we are unwilling to fully guarantee it."

D. Taking Away What Was Given: Internal Restrictions Within the CVRA

As the preceding discussion indicates, any assertion that the citizenry should expect protection from the government must be limited. Whatever initial glimmers of light the social contract theory[124] might cast on a claim for government protection, they are immediately dimmed by the constitutional limits articulated by the Supreme Court in the *DeShaney* and

---

121. MASS. GEN. LAWS. ch. 258B, § 3.

122. *Id.* at § 10.

123. Such an outright disclaimer of liability is not unusual. An equally potent example of this limited liability can be drawn from an examination of state and federal sexual offender registry programs. Many of these laws specifically recite that they were passed with the goal of "protecting the public." *See* FLA. STAT. ANN. § 755.21(3)(d) (2010); 42 U.S.C. § 14071(e)(2) (2006). While each program presents its own jurisdictional varieties, most require convicted sex offenders to register with an appropriate state or local agency, and then that agency is directed to release relevant information to the public regarding the registrant's residence. However, most of these statutes grant immunity to state actors for their good faith conduct. *See, e.g.*, FLA. STAT. ANN. § 775.21(9); 42 U.S.C. § 14971(f). For example, Ohio's law guides that except in cases of malicious, bad faith, or wanton or reckless behavior, state actors will be "immune from liability in a civil action to recover damages for injury, death, or loss to person or property allegedly caused by an act or omission in connection with a power, duty, responsibility, or authorization" under the state's sex offender registration laws. OHIO REV. CODE ANN. § 2950.12.

124. *See supra* Section II.A.

Gonzales decisions,[125] as well as by existing statutory torts and the common law.[126] In short, the promise of protection is highly ethereal, and enforced by courts in only the most select situations. Unfortunately, the protection right afforded to victims in the CVRA does little to overcome these problems.

The CVRA states that a victim has "[t]he right to be reasonably protected from the accused."[127] On its face, this language would appear to impose a statutory duty on the government to provide victims with protection. However, a closer examination of the statutory language and its legislative history reveals that the protection right is actually quite limited in its scope, and its enforceability is even narrower.

Even the framers of the CVRA, who in one breath used grand language regarding victim protection,[128] in the next breath acknowledged that the right is far narrower than suggested from a plain reading of the statute's terms. For example, in presenting the CVRA to Congress, Senator Kyl commented that: "Of course the government cannot protect the crime victim in all circumstances. However, where reasonable, the crime victim should be provided accommodations such as a secure waiting area, away from the defendant before and after and during breaks in the proceedings."[129] The Federal Attorney General Guidelines regarding victim services under the CVRA echo Senator Kyl's limiting language.[130] The Guidelines direct that, where possible, separate waiting areas should be provided to victims at trial or at parole hearings.[131] The Guidelines also note that victim protection services could also include "aiding a victim in changing his or her telephone number[,] to the extreme measure of proposing the victim for inclusion in the Federal Witness Security Program."[132] What is telling about the Attorney General Guidelines, is that

---

125. *See supra* Section II.B.

126. *See supra* Section II.C.

127. 18 U.S.C. § 3771(a)(1) (2006).

128. *See supra* notes 3–4 and accompanying text.

129. 150 CONG. REC. S10910 (daily ed. Oct. 9, 2004) (statement of Sen. Kyl); *see also* Kyl et al., *supra* note 3, at 596. As was brought to my attention by Professor Douglas Beloof, there are certainly grounds to contend that providing victims with a secure waiting area has value. Without the assurance of a secure waiting area, victims might be less willing to voluntarily participate in the criminal proceedings, and thereby might decline to exercise their other rights under the CVRA, such as the right to not to be excluded from the proceedings, and the right to be reasonably heard at plea, sentencing and parole proceedings. *See* 18 U.S.C. § 3771(a)(2)–(3).

130. *See* OFFICE FOR VICTIMS OF CRIME, OFFICE OF JUSTICE PROGRAMS, U.S. DEP'T OF JUSTICE, ATT'Y GEN. GUIDELINES FOR VICTIM AND WITNESS ASSISTANCE 30–32, 35 (2005).

131. *Id.*

132. *Id.* at 25. One must question, whether, except for the most extreme of circumstances, crime victims would consider participation in the Federal Witness Security Program a desirable right afforded under the CVRA, much less an appropriate response to their victimization.

they make clear that the Guidelines should in no way be construed to "require personal protection of a victim, such as by bodyguards."[133] Finally, and perhaps hinting at the "reasonableness" qualification that appears in the statute, the legislative history of the CVRA suggests that the victim's right to be reasonably protected from the accused includes considering the victim's safety in the course of court determinations regarding a defendant's release pending or during trial.[134] As noted by Senator Kyl, "the right to protection also extends to require reasonable conditions of pre-trial and post-conviction [release] that include protections for the victim's safety."[135]

Legislative intent and administrative materials analyzing the CVRA suggest that certain actions can be taken on behalf of victims, which might serve as a means to afford them safety. At one end of the spectrum, the CVRA could provide protection to victims by giving them separate waiting areas during the defendant's trial or at a parole hearing.[136] At the other end of the spectrum, victims could enroll in a witness protection program.[137] And somewhere in the middle, there is the hint that the right to reasonable protection is related to decisions about the defendant's release or parole.[138] As I will discuss later, it is this latter formulation that I believe gives greatest meaning to the victim's right to reasonable protection from the accused. However, regardless of how one interprets the current meaning of the CVRA's protection language, the statute severely limits a victim's ability to enforce her protection right.

First, the CVRA uses largely deferential language in describing the duties the statute imposes on the courts and Department of Justice employees. Courts are imposed with the seemingly mandatory command that they "shall ensure" that crime victims are afforded their rights.[139]

---

133. *Id.*

134. *See* 150 CONG. REC. S10910 (daily ed. Oct. 9, 2004) (statement of Sen. Kyl).

135. 150 CONG. REC. S10910 (daily ed. Oct. 9, 2004) (statement of Sen. Kyl); *see also* Kyl et al., *supra* note 3, at 596.

136. *See* 150 CONG. REC. S10910 (daily ed. Oct. 9, 2004) (statement of Sen. Kyl); OFFICE FOR VICTIMS OF CRIME, supra note 130, at 35.

137. OFFICE FOR VICTIMS OF CRIME, supra note 130, at 25.

138. *See* Kyl et al., *supra* note 3, at 601–09.

139. 18 U.S.C. § 3771(b)(1). As a matter of statutory construction, the word "shall" is generally interpreted as mandating certain action, while the word "may" is permissive in nature. *See, e.g.*, Alatech Healthcare, L.L.C. v. United States, 89 Fed. Cl. 750, 753 (Fed. Cl. 2009) (quoting BLACKS LAW DICTIONARY); LeMay v. United States Postal Service, 450 F.3d 797, 799 (8th Cir. 2006); Keith v. Rizzuto, 212 F.3d 1190, 1193 n.3 (10th Cir. 2000); *see also* 82 C.J.S. *Statutes* § 498 (2009). However, the Supreme Court was unwilling to employ this generally accepted approach to statutory construction when it interpreted the mandatory language appearing in Jessica Gonzales' restraining order and the identical language appearing in state statute. *See supra* notes 35–51 and accompanying text. Nonetheless, since the passage of the CVRA, it appears that there have been several courts that have taken the statute's mandatory language to heart and have taken steps to ensure that victims are

Prosecutors are held to a lesser standard, being directed that they, along with other Department of Justice employees, shall use their "best efforts" to accord victims their rights.[140] Despite the mandatory "shall ensure" language imposed on the courts, and the more relaxed "best efforts" language directed to other government employees, the CVRA nonetheless relieves all such actors from any individual damage liability for failure to comply with the statute's terms. The statute further explicitly disclaims that it creates or imposes any specific duties on behalf of the United States or its officers, the breach of which could result in a claim for damages.[141] At most, employees who willfully or wantonly fail to comply with provisions of the law could be disciplined.[142] Therefore, regardless of what promise of protection the CVRA makes to victims, it is wholly undermined by the statute's disclaimer of any actionable duty imposed on government employees to fulfill the law's terms.

Second, to the extent that the CVRA grants victims a means to enforce their rights under the CVRA, their remedies are limited. Under the statute's enforcement terms, victims are permitted to submit a motion for relief to the district court in which the defendant is being prosecuted or where the crime occurred.[143] If the court denies the victim's motion for relief, the victim is permitted to petition to the relevant court of appeals with a writ of

---

provided their rights. *See* United States v. Rubin, 558 F. Supp. 2d. 411, 428–29 (E.D.N.Y. 2008); United States v. Degenhardt, 405 F. Supp. 2d 1341, 1343 (D. Utah 2005); United States v. Turner, 367 F. Supp .2d 319, 323–24 (E.D.N.Y. 2005).

140. § 3771(c)(1).

141. *Id.* The statute reads:

> Nothing in this chapter shall be construed to authorize a cause of action for damages or to create, to enlarge, or to imply any duty or obligation to any victim or other person for the breach of which the United States or any of its officers or employees could be held liable in damages.

Id.

142. 28 C.F.R. § 45.10(e)(1) (2009). Specifically, the regulation guides that:

> If, based on an investigation, the [Victims' Rights Ombudsman] VRO determines that a Department of Justice employee has wantonly or willfully failed to provide the [victim] with a right listed in [the CVRA], the VRO shall recommend, in conformity with laws and regulations regarding employee discipline, a range of disciplinary sanctions to the head of the Office of the Department of Justice in which the employee is located, or to the official who has been designated by the Department of Justice regulations and procedures to take action on disciplinary matters for that office.

Id. Government employees who fail to provide victims their rights, but do so in a manner that is not willful or wanton manner, must undergo additional training on victims' rights. *See Id.* at § 45.10(d).

143. 18 U.S.C. § 3771(d)(3).

mandamus.[144] Even if the appellate court finds that the victim's rights were violated, the victim's remedies are limited to re-opening a plea or sentencing hearing, and then only if a variety of additional statutory requirements are satisfied.[145] Therefore, if a victim believes she has been denied her "right to be reasonably heard at any public proceeding in the district court involving . . . plea [or] sentencing,"[146] then the CVRA affords her some relief. However, if the victim believes she has been denied any of her other rights under the statute,[147] including the right to be reasonably protected from the accused, she has no specific means to seek redress for those violations.

The CVRA's protection right presents many of the same problems that arise with the state-based protection rights granted to victims.[148] While the terms of the protection right could be read broadly, upon further investigation, it becomes obvious that the promise of protection is by no means clear, and its enforcement precarious. At a minimum, the right promises something that cannot be fulfilled, thereby creating unrealistic expectations in victims regarding what they can expect from the criminal justice system. One can only assume that Jessica Gonzales and Catherine Ford were sorely confused and disappointed to learn that the promises of

---

144. *Id.*

145. *Id.* § 3771(d)(5). Specifically, the statute states:

> A victim may make a motion to re-open a plea or sentence only if (A) the victim has asserted the right to be heard before or during the proceeding at issue and such right was denied; (B) the victim petitions the court of appeals for a writ of mandamus within 10 days [of the denial of the right]; and (C) in the case of a plea, the accused has not pled to the highest offense charged.

Id.

146. *Id.* § 3771(a)(4).

147. The other rights afforded to victims under the CVRA include:

> [t]he right to reasonable, accurate, and timely notice of any public court proceeding, or any parole proceeding, involving the crime or of any release or escape of the accused; [t]he right not to be excluded from any such public court proceeding, unless the court, after receiving clear and convincing evidence, determines that testimony by the victim would be materially altered if the victim heard other testimony at that proceeding; . . . [t]he reasonable right to confer with the attorney for the Government in the case; [t]he right to full and timely restitution as provided in law; [t]he right to proceedings free from unreasonable delay; [and t]he right to be treated with fairness and with respect for the victim's dignity and privacy.

Id. at §§ 3771(a)(2), (3), (5)–(8).

148. *See supra* Section II.C.

protection upon which they relied meant very little.[149] Such unfulfilled promises are problematic and should give victims' rights advocates pause.

When victims' rights laws make promises to victims that cannot be fulfilled, the very effectiveness, legitimacy and power of those laws are undermined. Most victims enter the criminal justice system traumatized and untrusting as a result of their victimization.[150] Making promises to victims that can quickly and easily evaporate does little to bolster victim participation or trust in the laws passed for their benefit.[151] Hence, a promise of protection may cause more harm to victims than it does to help them.

However, the limitations that currently undermine the victim's protection right are redeemable. First, the other rights granted to victims under the CVRA, and the means by which victims can enforce those rights, provide a model for how to approach the victim's protection right.[152] Second, federal law provides analogous examples of how victim safety and protection concerns can be addressed.[153] These alternative examples are best explained through procedural justice theory, and the victim's protection right should also be viewed in this context.[154]

In order to best understand how procedural justice theory can provide more substance to the victim's protection right, we must step back and engage in a broader view of how the criminal justice system has incorporated, and currently incorporates, the victim in the prosecutorial process. The ensuing discussion will be both historical and theoretical, charting the rise and fall of the victim within the criminal justice system, the theoretical constructs that have influenced the ebb and flow of victim prominence, and the eventual emergence of the victims' rights movement. Out of this historical and theoretical review, I conclude that procedural justice theory provides the best explanation for why victims and their interests should be incorporated into standard criminal processes. Therefore, when crafting victims' rights laws, legislators' should treat as paramount procedural justice principles. In so doing, procedural justice theory can bring meaning and enforceability to the victim's CVRA protection right.

---

149. *See supra* note 50 and accompanying text regarding *Town of Castle Rock v. Gonzales*; *supra* note 110 and accompanying text regarding *Ford v. Town of Grafton*.

150. *See infra* note 212 and accompanying text.

151. *See infra* notes 212–216 and accompanying text.

152. *See* 18 U.S.C. § 3771 (2006).

153. *See infra* notes 327–334 and accompanying text.

154. *See infra* Section III.

III. A THEORETICAL REVIEW: THE VICTIM IN THE CRIMINAL PROCESS

A. The Rise and Fall of Victim Primacy

In our earliest criminal justice systems, the victim was responsible for holding the perpetrator accountable for his wrongful acts. In the closely-knit fabric of tribal and family-based cultures, individual victim responses to criminal wrongdoing established the norms for appropriate retaliation and compensation for criminal deeds.[155] As these early systems developed, the offender was often required to provide direct restitution to the victim for the harm she suffered.[156] Therefore, it made sense that the victim was the primary party vested with the power to bring actions against the liable party. If the person who would benefit most from the proceeding was the victim, then the victim should possess a primary role in overseeing and controlling that event.[157]

A victim-centered private prosecution model dominated for much of our nation's early formation.[158] However, around the time of the American Revolution, the focus and structure of the American criminal justice system began to evolve.[159] Crime shifted from being viewed as an injury suffered by a discrete individual to a violation "which tears at the fabric of our peace and community."[160] Philosophers and governmental leaders stopped viewing crime as a personal and isolated episode between a victim and offender, and instead as an event that implicated broader concerns of how we relate to and function with one another as a community.[161] Essentially, crime represented a breach of the social contract.[162] The core reason, then, for a criminal justice system was to serve the interests of society by consistently and systematically holding accountable those who breached the

---

155. Peggy M. Tobolowsky, *Victim Participation in the Criminal Justice Process: Fifteen Years After the President's Task Force on Victims of Crime*, 25 NEW. ENG. J. ON CRIM. & CIV. CONFINEMENT 21, 23 (1999).

156. *See* Cellini, *supra* note 16, at 841 (referencing the requirement of restitution to victims as appearing in Code of Hammurabi, the Old Testament, Greek and Roman penal codes, and early Anglo-Saxon law); McDonald, *supra* note 11, at 652–53 ("a system of restitution by the offender to the victim . . . was an accepted goal of the system.").

157. *See* Cellini, *supra* note 16, at 841.

158. *See generally* Douglas E. Beloof & Paul G. Cassell, *The Crime Victim's Right to Attend the Trial: The Reascendant National Consensus*, 9 LEWIS & CLARK L. REV. 481, 484–85 (2005); Douglas E. Beloof, *Weighing Crime Victims' Interests in Judicially Crafted Criminal Procedure*, 56 CATH. U. L. REV. 1135, 1138–39 (2007) [hereinafter *Weighing Crime Victims' Interests*]; McDonald, *supra* note 11, at 651–654; Tobolowsky, *supra* note 155, at 25; Cellini, *supra* note 16, at 842–43.

159. *See* Cellini, *supra* note 16, at 844–47.

160. Steven J. Twist, *The Crime Victims' Rights Amendment and Two Good and Perfect Things*, 1999 UTAH L. REV. 369, 369 (1999).

161. *See* Tobolowsky, *supra* note 155, at 26.

162. *See supra* notes 16–19 and accompanying text.

social contract. Therefore, it was inappropriate to allow individual victims, and their potential desires for revenge and vengeance, to control the process.[163] A criminal justice system that delegates to victims the task of seeking justice could result in unmanageable blood feuds and undermine ordered society.[164] Instead, a state sanctioned system would more likely assure fair and efficient prosecutions.

It is folly, however, to argue that criminal prosecutions should be entirely cleansed of the human desire for vengeance or revenge: "[T]he basic urge for [revenge] is a cultural universal, across time and place, and it establishes itself early in life."[165] While there are legitimate reasons to temper individual and unfettered acts of vengeance, "most typical, decent, mentally healthy people have a kind of commonsense approval of some righteous hatred and revenge."[166] Nor should revenge be viewed entirely as a destructive emotion to be eliminated and banished from the law. In fact, some scholars have contended that revenge should be viewed as "an ennobling, as well as an enabling, concept."[167] One scholar points out:

> Revenge cultures encompassed a sense of self worth; that is, the recognition that no one had the right to inflict unprovoked harm upon another, and that when another did so, it was the victim, and not the community at large, who had primarily been wronged. As such, the victim had the right (or in some cultures the duty), to personally recapture his respect and honor. In effect, cultures which permitted revenge, allowed those victims strong enough (or from families with sufficient strength) to erase the psychological stigma of their victimhood.[168]

---

163. McDonald, *supra* note 11, at 652, 655–66; *see also* Cellini, *supra* note 16, at 847–56 (discussing a time line of victims' rights and the development of the law up to present day); Tobolowsky, *supra* note 155, at 25–26 (discussing reasons for shifting the control of the process).

164. Kenworthy Bliz, *The Puzzle of Delegated Revenge*, 87 B.U. L. REV. 1059, 1072 (2007); *see also* Steven Eisenstat, *Revenge, Justice and Law: Recognizing the Victim's Desire for Vengeance as a Justification for Punishment*, 50 WAYNE L. REV. 1115, 1117–18 (2004) (addressing concerns of disproportionate results which could arise out of "blood feuds").

165. Bliz, *supra* note 164, at 1063. Professor Bliz notes that studies even indicate that non-human primates—chimpanzees—experience feelings of revenge and retribution. *Id.* (citing Frans B.M. de Waal, *The Chimpanzee's Sense of Social Regularity and Its Relation to the Human Sense of Justice*, *in* THE SENSE OF JUSTICE 241, 241–45 (Roger D. Masters & Margaret Gruter eds., 1992); Toshisada Nishida et al., *A Within-Group Gang Attack on a Young Adult Male Chimpanzee: Ostracism of an Ill-Mannered Member?*, 36 PRIMATES 207, 209–10 (1995)).

166. Jeffrie J. Murphy, *Getting Even: The Role of the Victim*, *in* RETRIBUTION RECONSIDERED: MORE ESSAYS IN THE PHILOSOPHY OF LAW 65, 65 (1992).

167. *See* Eisenstat, *supra* note 164 at 1148.

168. *Id.* at 1149.

Therefore, a legal system which acknowledges victim vengeance, or at the very least, the victim's desire to see justice, should not, in and of itself, be viewed as faulty. The more relevant concern is how a criminal justice system should effectively marshal these core human emotions.

B. The Public Prosecution Model

Most legal theorists have concluded that a state-controlled public prosecution model is the appropriate means by which to impliedly acknowledge the individual victim's desire for justice, while also emphasizing the social contract ideal that crime represents a wrong committed against all of society and requires a formal, state-sanctioned response.[169] Accordingly, under the development of the public prosecution model, criminal trials shifted from being victim-controlled events to being public affairs, which were initiated, overseen, and marshaled by professional prosecutors.[170] As a result of this philosophical shift, however, the individual harm suffered by the victim was transformed into a proxy of the harm suffered by the collective state.[171] Criminal prosecutions increasingly focused on balancing society's interests in maintaining safety and civil order against the defendant's liberty interest, and less on any interests the victims might have had in the proceeding.[172] Victim centered restitution measures fell to the wayside, and incarceration became the dominant form of punishment.[173]

Two competing value systems have shaped the public prosecution model and contributed to diminishing the victim's place within the criminal justice system. As originally articulated by Professor Herbert L. Packer, the Crime Control and Due Process Models served to collectively assure that the criminal justice system furthered the goals of the social contract.[174] The Crime Control Model is "based on the proposition that the repression of criminal conduct is by far the most important function to be performed by the criminal process."[175] This model focuses on efficient crime control, and suggests that governments prosecute crime to further the broader interests

---

169. *See, e.g.*, Cellini, *supra* note 16, at 843–44; Tobolowsky, *supra* note 155, at 103.

170. Cellini, *supra* note 16, at 844.

171. *Id.* at 845.

172. *Id.* at 846–47.

173. *Id.* at 847–48; Tobolowsky, *supra* note 155, at 26.

174. *See* Herbert L. Packer, *Two Models of the Criminal Process*, 113 U. PA. L. REV. 1, 9 (1964–65).

175. *Id.*; *see also* Douglas E. Beloof, *The Third Model of Criminal Process: The Victim Participation Model*, 1999 UTAH L. REV. 289, 290–93 (1999), [hereinafter *The Third Model*]; John Griffiths, *Ideology in Criminal Procedure or A Third "Model" of the Criminal Process*, 79 YALE L. J. 359, 360–67 (1969–70) (discussing Professor Packer's crime models); Kent Roach, *Four Models of the Criminal Process*, 89 J. CRIM. L. & CRIMINOLOGY 671, 676–80 (1999).

of society, rather than the interests of individual crime victims.[176] As stated by Professor Packer, "[t]he failure of law enforcement to bring criminal conduct under tight control is viewed as leading to the breakdown of public order and thence to the disappearance of an important condition of human freedom."[177] The system's swift response to criminal behavior serves to correct the offender's breach of the social contract, and also serves to deter those contemplating future breaches of the social contract.[178] The Crime Control Model also serves to maintain the citizenry's trust that their government will fulfill the task of holding accountable those who have violated the contract.[179] Through the swift and efficient response to crime, the government sends an important message to the governed: "We acknowledge your harm and seek to eliminate any further harm."

The Crime Control Model aligns nicely with utilitarian justifications for our criminal justice system. Utilitarian theory broadly states that the primary goal of any moral system is to maximize happiness.[180] In the context of criminal law, utilitarianism posits that there are three primary justifications for why we prosecute and punish offenders. First, prosecution and punishment is justified to deter the current offender from committing future crimes (specific deterrence) as well as to deter members of the public from committing similar wrongs (general deterrence).[181] Second, prosecutions and subsequent incarcerations are justified as a means to incapacitate the offender from committing additional crimes.[182] Finally, a utilitarian approach to prosecution and punishment asserts that the

---

176. *See* Packer, *supra* note 174, at 10–11.

177. *Id.* at 9.

178. *See id.* at 10–11.

179. *See* Roach, *supra* note 175, at 677.

180. In discussing utilitarianism in the context of criminal law, the concept is also characterized as consequentialism, or corrective justice. *See* Russell L. Christopher, *Deterring Retributivism: The Injustice of "Just" Punishment*, 96 NW. U. L. REV. 843, 857 (2002). *See generally* JEREMY BENTHAM, THEORY OF LEGISLATION 1–4 (C.K. Ogden, ed., 1931) JOHN STUART MILL, UTILITARIANISM, LIBERTY, AND REPRESENTATIVE GOVERNMENT 5–24 (1910); Eisenstat, *supra* note 164, at 1127; Richard R. Frase, *Punishment Purposes*, 58 STAN. L. REV. 67, 70–75 (2005); Matthew Haist, *Deterrence in a Sea of "Just Deserts:" Are Utilitarian Goals Achievable in a World of "Limiting Retributivism?"*, 99 J. CRIM L. & CRIMINOLOGY 789, 793–99, 817–21 (2009); Andrew R. Strauss, *Losing Sight of the Utilitarian Forest for the Retributivist Trees: An Analysis of the Role of Public Opinion in a Utilitarian Model of Punishment*, 23 CARDOZO L. REV. 1549 (2002) (discussing utilitarian theory and criminal law).

181. Christopher, *supra* note 180, at 857.

182. *Id.*; *see also* Linda S. Beres & Thomas D. Griffith, *Do Three Strikes Laws Make Sense? Habitual Offender Statutes and Criminal Incapacitation*, 87 GEO. L.J. 103, 112 (1998); Markus Dirk Dubber, Note, *The Unprincipled Punishment of Repeat Offenders: A Critique of California's Habitual Criminal Statute*, 43 STAN. L. REV. 193, 215–23 (1990) (discussing incapacitation theory); Andrew D. Leipold, *Recidivism, Incapacitation, and Criminal Sentencing Policy*, 3 ST. THOMAS L.J. 536, 541–44 (2005).

rehabilitation of the offender will result in an overall social good.[183] The Crime Control Model's goal to repress criminal conduct and maintain social order fulfills these utilitarian ideals.

Conversely, the Due Process Model focuses on "the reliability of fact finding processes."[184] If, according to Professor Packer, the Crime Control Model could be likened to an assembly line down "which moves an endless stream of cases,"[185] the Due Process Model is akin to an "obstacle course."[186] While the desirability of repressing crime is not disregarded under the Due Process Model, it emphasizes that in prosecuting crime, the "primacy of the individual [defendant] and the complementary concept of limitation on official power" should be central.[187] The Due Process Model seeks to prevent and eliminate any mistakes the state might make during the adjudicative process that could result in the government's abuse of its prosecutorial power.[188]

By counterbalancing the Crime Control Model with the Due Process Model, the goals of the social contract can be fulfilled in a holistic manner.[189] In departing from a state of nature and creating structured societies, citizens ceded to the government their individual ability to protect themselves along with the right to fulfill any private vengeance against wrongdoers.[190] The Crime Control Model assures that crime will be addressed swiftly and efficiently, thereby restoring the social order undermined by the criminal's acts. However, in order to ensure that the state does not abuse the prosecutorial power granted to it by the citizenry, the Due Process Model imposes limits on the Crime Control Model.[191] "The aim of the process is at least as much to protect the factually innocent as it is to convict the factually guilty."[192] In this regard, the Due Process Model can be aligned with a retributive approach to criminal justice theory.[193]

---

183. Christopher, *supra* note 180, at 857.

184. Packer, *supra* note 174, at 14.

185. *Id.* at 11.

186. *Id.* at 13.

187. *Id.* at 16.

188. *Id.*

189. *See* Cellini, *supra* note 16, at 841.

190. Packer, *supra* note 174, at 10.

191. *See id.* at 13.

192. *Id.* at 15.

193. Philosophers tend to argue that one must be either a utilitarian or retributivist when discussing what, how, and why we acknowledge and punish crime. *See, e.g.*, Christopher, *supra* note 180, at 845–55; Michael Moore, *Victims and Retribution: A Reply to Professor Fletcher*, 3 BUFF. CRIM. L. REV. 65, 66–67 (1999). *Cf.* Benjamin B. Sendor, *Restorative Retributivism*, 5 J. CONTEMP. LEGAL ISSUES 323, 335 (1994) (adopting a "pluralist" approach to criminal theory). Conversely, lawmakers tend to be far more willing to coalesce the theories. *See, e.g.*, 18 U.S.C. § 3553(a)(2) (2006) (showing that United States Sentencing Guidelines combine both utilitarian and retributive justifications for calculating a defendant's sentence).

Retributive theory is generally rooted in a concern that any official response to criminal behavior should be tempered and controlled, rather than represent an unadulterated expression of victim revenge.[194] Retributivists contend that in order to justify the criminal justice system and its resulting punishments, the offender's punishment must be limited to that which he exactly deserves. In this regard, retributivists invoke Immanuel Kant's maxim that each should "[a]ct so that you treat humanity, whether in your own person or in that of another, always as an end and never as a means only."[195] According to general retributivist theory, criminal prosecutions and their resulting punishment should ensure that the offender bears the burden of his unlawful acts only in direct proportion to the harm he caused. The punishment must fit the crime, no more and no less.[196] Should the punishment advance any other purpose, such as deterring future crime or rehabilitating the offender, retributivists argue the criminal justice system runs afoul of Kant's maxim that we should not treat one another as means to an end.[197]

When examining how these theories may help further advance the victim's role in the criminal justice system, and more specifically, give value to the victim's right of protection from the accused, I place myself in the latter camp. The extensive body of legal and philosophical scholarship makes clear that there likely cannot be one unified or perfect theory for why we prosecute crimes or why we should (or should not) seek to more fully integrate victims into the criminal process. *See generally* Frase, *supra* note 180; Haist, *supra* note 180. However, these divergent and sometimes conflicting theories nonetheless help better inform the conversation of "why," and hence should not be discarded merely because one or the other cannot perfectly respond to all the challenges posed against them.

194. There is a vast body of literature regarding retributive justifications for the criminal justice system. A quick review of this literature makes clear that even among retributivist scholars, clearly defining retributivism can be a tricky task. *See, e.g.*, Christopher, *supra* note 180, at 865–67 (noting different variations of retributivism).

195. A.I. Melden, *Dignity, Worth, and Rights*, *in* THE CONSTITUTION OF RIGHTS: HUMAN DIGNITY AND AMERICAN VALUES 29 (Michael J. Meyer & William A. Parent eds. Cornell U. Press 1992) (quoting Immanuel Kant, THE FOUNDATIONS OF THE METAPHYSICS OF ETHICS (Lewis White Beck, trans. 1959)).

196. *See* Christopher, *supra* note 180, at 848. It is here where a retributive approach to criminal theory dovetails with Professor Packer's Due Process Model. *See supra* notes 184–187 and accompanying text. To the extent that retributive theory seeks to ensure that only the guilty be punished, and only to the extent which they deserve, one can draw a parallel with Professor's Packer's Due Process Model, in that when prosecuting crimes, the "primacy of the individual [defendant] and the complementary concept of limitation on official power" must be acknowledged. Packer, *supra* note 174, at 16.

197. In this regard, retributivists tend to claim their approach to criminal punishment is superior to a utilitarian approach. *See* Christopher, *supra* note 180, at 933. However, Professor Russell Christopher raises a compelling argument that retributivists may be just as guilty as utilitarians in treating individuals as a means to an end. *Id.* In contrast to assertions that utilitarianism uses offenders as a means to achieve positive ends, such as reduced crime rates or a safer community, Professor Christopher accuses retributive theory of using victims as a means to the end—determining the just punishment of the offender. Christopher, *supra*

Similarly, traditional retributive theory rejects any argument that the criminal justice system should be attuned to victim interests. Shifting our attention from "what does the offender deserve?" to "what does the victim want?" would become immediately utilitarian. For example, an overly vengeful victim may desire an inappropriately harsh sanction for the offender, while conversely, a particularly forgiving victim may not desire any punishment at all.[198] Such a state of affairs would undermine the retributivist goal of ensuring that punishment is imposed based only on desert.[199] Professor Packer's Due Process Model furthers this retributive goal by imposing throughout the prosecutorial process certain "quality control" checks to ensure that only those genuinely worthy of punishment receive their just deserts.[200]

The public prosecution model, shaped by Professor Packer's theories, strives to ensure that crimes are dealt with swiftly and efficiently, but that defendants are also treated fairly and are not subject to unfettered victim desires for revenge. However, by focusing on the balance between the desires of society and the rights of the defendant, the public prosecution model pushed victims to the sidelines.[201] Victims could report the crime and provide evidence when called upon by the state, but were otherwise expected to "behave like Victorian children—seen but not heard."[202] Even more harshly put,

> [s]ince crime was conceptualized as an event that threatened and offended the entire community, and was prosecuted by the state on behalf of an abstraction (i.e. "the People"), the real flesh-and-blood victim was treated

---

note 180, at 933–53.

198. Criminal procedure has nonetheless evolved to be far more solicitous to hearing from the victim on the matter of sentencing in the form of victim impact statements. *See* Payne v. Tennessee, 501 U.S. 808, 821–24 (1991); United States v. Blake, 89 F. Supp. 2d 328, 349–53 (E.D.N.Y. 2000). However, victim impact statements are not without controversy and continue to be met with significant academic debate. *See generally* Susan Bandes, *Empathy, Narrative, and Victim Impact Statements*, 63 U. CHI. L. REV. 361 (1996); Paul G. Cassell, *In Defense of Victim Impact Statements*, 6 OHIO ST. J. CRIM. L. 611 (2009); Justin D. Flamm, *Due Process on the "Uncharted Seas of Irrelevance": Limiting the Presence of Victim Impact Evidence at Capital Sentencing After* Payne v. Tennessee, 56 WASH. & LEE L. REV. 295 (1999); Joe Frankel, *Payne, Victim Impact Statements, And Nearly Two Decades of Devolving Standards of Decency*, 12 N.Y. CITY L. REV. 87 (2008); Trey Hill, *Victim Impact Statements: A Modified Perspective*, 29 LAW & PSYCHOL. REV. 211 (2005); Jody Lyneé Madeira, *"Why Rebottle the Genie?": Capitalizing on Closure in Death Penalty Proceedings*, 85 IND. L.J. 1477 (2010); Beth E. Sullivan, *Harnessing* Payne: *Controlling the Admission of Victim Impact Statements to Safeguard Capital Sentencing Hearings From Passion and Prejudice*, 25 FORDHAM URB. L.J. 601 (1998).

199. *See* Christopher, *supra* note 180, at 933.

200. Packer, *supra* note 174, at 14–15.

201. Cellini, *supra* note 16, at 847–48; McDonald, *supra* note 11, at 655–56.

202. Kenna v. U.S. District Court, 435 F.3d 1011, 1013 (9th Cir. 2006).

> just like any other piece of evidence, a mere exhibit to be discarded after the trial.[203]

Thus, in direct contradiction to Kant's edict, the victim was rendered a mere means to an end by which the state could fulfill its interest in holding accountable those who had breached the social contract.[204]

The public prosecution model also stripped victims of any legal or enforceable interest they may have otherwise held in an action against the offender. This reality was highlighted in the Supreme Court's ruling in *Linda R.S. v. Richard D.*[205] In *Linda R.S.*, a single mother brought an action challenging the state prosecutor's failure to bring a criminal action against the father of her child for not paying child support.[206] The state prosecutor declined to bring such an action, as the state had consistently interpreted the child support statute as applying only to the parents of legitimate children and not to the parents of children born out of wedlock.[207] The Supreme Court rejected the mother's case against the prosecutor, holding that she lacked standing.[208] According to the court, even if the mother and her child were indeed victims of the father's failure to pay child support, the mother did not possess a legally recognized interest to compel any action by the government on her behalf. The Court reasoned that "in American jurisprudence . . . a private citizen lacks a judicially cognizable interest in the prosecution or nonprosecution of another."[209] Linda R.S. might have been harmed, but under a public prosecution framework, the lack of official response to her harm was not something about which she had legal grounds to complain.[210] At most, she was reporting as a witness, rather than someone with a direct interest in the case and its outcome.

---

203. Andrew J. Karmen, *Who's Against Victims' Rights? The Nature of the Opposition to Pro-vicitm Initiative in Criminal Justice*, 8 ST. JOHN'S J. LEGAL COMMENT. 157, 158 (1992); *see also* Aynes, *supra* note 5, at 68 (suggesting that victims are not so much forgotten by the criminal justice system, but rather used by it).

204. McDonald, *supra* note 11, at 650 ("In contemporary criminal justice the victim serves only as a means to an end, namely, a piece of evidence to be used by the state to obtain a conviction. The only concern that the state has for the victim is his willingness to cooperate and his ability to be a convincing witness."); *see also* Christopher, *supra* note 180, at 851 (discussing how retributive theory treats victims as a means to an end).

205. 410 U.S. 614 (1973); *see also* Beloof, *Weighing Crime Victims' Interests*, *supra* note 158 at 1141–44 (discussing the *Linda R.S.* case).

206. *Linda R.S.*, 410 U.S. at 615–16.

207. *Id.*

208. *Id.* at 619.

209. *Id.* The Court did note, however, that "Congress may enact statutes creating legal rights, the invasion of which creates standing, even though no injury would exist without the statute." *Id.* at 617 n.3.

210. Under the public prosecution framework, the prosecutor is also given a great deal of discretion to decide whether and under what circumstances to bring criminal charges. *See, e.g.*, United States v. Batchelder, 442 U.S. 114, 124–26 (1979); Hardwick v. Doolittle, 558

The public prosecution model, therefore, moored in its utilitarian and retributive foundations, largely disregards the interests victims have in criminal prosecutions. Being a victim of crime does not create a right to see that offense prosecuted, or a right to see the offender brought to justice. However, victims have railed against their diminished status. Over the last thirty years, they have reshaped how the criminal justice system responds to, and integrates victims into the criminal process.

C. Victims Respond: The Emergence of the Victims' Rights Movement

Victim dissatisfaction with their treatment under the public prosecution model spurned the victims' rights movement.[211] Of course, the direct and immediate impact of crime tends to leave victims awash in feelings of powerlessness, isolation, shame, anger, and fear.[212] However, many studies indicate that victims are often more affected by their treatment throughout the course of their limited involvement in the prosecutorial process than by the crime itself.[213] As noted by the framers of the CVRA,

> [i]n case after case we found [that] victims, and their families, were ignored, cast aside, and treated as non-participants in a critical event in their lives. They were kept in the dark by prosecutors to [sic] busy to care enough, by judges focused on defendant's rights, and by a court system that simply did not have a place for them.[214]

Sociological studies have further indicated that as a result of this so called "secondary victimization," victims' self-esteem, faith in the future, trust in the legal system, and belief in a just world are undermined to depths beyond

---

F.2d 292, 301–303 (5th Cir. 1977); State v. Conger, No. 2008AP755-CR, 2010 WL 2595337, at *5–6 (Wis. 2010); Barbara O'Brien, *A Recipe for Bias: An Empirical Look at the Interplay Between Institutional Incentives and Bounded Rationality in Prosecutorial Decision Making*, 74 MO. L. REV. 999, 1005–07 (2009); Michael A. Simons, *Prosecutors as Punishment Theorists: Seeking Sentencing Justice*, 16 GEO. MASON L. REV. 303, 307–10 (2009); Evangeline A. Zimmerman, *The Federal Sentencing Guidelines: A Misplaced Trust in Mechanical Justice*, 43 U. MICH. J.L. REFORM 841, 852 (2010).

211. *See generally* David E. Aaronson, *New Rights and Remedies: The Federal Crime Victims' Rights Act of 2004*, 28 PACE L. REV. 623 (2008); Beloof, *Weighing Crime Victims' Interests*, *supra* note 158; Paul G. Cassell, *Treating Crime Victims Fairly: Integrating Victims into the Federal Rules of Criminal Procedure*, 2007 UTAH L. REV. 861, 865–70 (2007); Kyl et al., *supra* note 3; Tobolowsky, *supra* note 155.

212. Ilyssa Wellikoff, *Victim-Offender Mediation and Violent Crimes: On the Way to Justice*, 5 CARDOZO J. CONFLICT. RESOL. 2, 8 (2003).

213. *See* Edna Erez & Pamela Tontodonato, *Victim Participation in Sentencing and Satisfaction with Justice*, 9 JUST. Q. 393, 394 (1992).

214. *Senate Floor Statements in Support of the Crime Victim's Rights* Act, 19 FED. SENT'G REP. 62, 63 (October 2006) (statement of California Senator Dianne Feinstein in support of S. 2329, 108th Congress (2004)).

any personal suffering resulting from the crime itself.[215] One victim stated that her "sense of disillusionment with the judicial system is many times more painful [than the crime itself]. I could not, in good faith, urge anyone to participate in this hellish process."[216] In response to this overwhelming victim dissatisfaction, both individual states and the federal government began passing victims' rights laws. Over the last thirty years, every state has enacted some form of victims' rights legislation and nearly two-thirds have passed amendments to their state constitutions granting victims' rights in the criminal justice process.[217] The same is true on the federal level. Since the early 1980s, Congress has passed a series of progressively effective victims' rights laws, the most recent being the CVRA.[218]

Victims' rights advocates tend to advance two core arguments for why victims' rights laws are necessary. First, advocates have argued that the justice system needs to do a better job at acknowledging victims' interests in criminal proceedings.[219] Even within the context of the public prosecution model, an individual has suffered harm and that harm should be acknowledged.[220] As aptly expressed by one victim, "'The State of New

---

215. *See* Marilyn Peterson Armour & Mark S. Umbreit, *The Ultimate Penal Sanction and "Closure" for Survivors of Homicide Victims*, 91 MARQ. L. REV. 381, 415 (2007); *see also* Erez & Tontodonato, *supra* note 213, at 393–94 (noting prevailing victim dissatisfaction with treatment by the criminal justice system); Uli Orth, *Secondary Victimization of Crime Victims by Criminal Proceedings*, 15 SOC. JUST. RES. 313, 314 (2002) (noting additional psychological harm suffered by victims as a result of contact with the criminal justice system); Tobolowsky, *supra* note 155, at 27 (noting studies examining "the psychological impact of victimization on victims, as well as the impact of their significant exclusion from the criminal justice process."); Pamela Tontodonato & Edna Erez, *Crime, Punishment and Victim Distress*, 1994 INT. REV. VICTIMOLOGY 33, 34–36 (1994) (examining the effect of victim participation in the prosecutorial process on victim distress).

216. S. Rep. No. 97-532, at 537 (1982) *reprinted in* 1982 U.S.C.C.A.N. 2515, 2543.

217. *See, e.g.*, ALA. CONST. art. I. § 6.01; ALASKA CONST. art. I, § 24; ARIZ. CONST. art. 2, § 2.1; CAL. CONST. art. 1, § 28(a)–(b); COLO. CONST. art. II, § 16a; CONN. CONST. art. XXIX; FLA. CONST. art. 1, § 16(b); IDAHO CONST. art. I, § 22; ILL. CONST. art. 1, § 8.1; IND. CONST. art. I, § 13(b); KAN. CONST. art. 15, § 15; LA. CONST. art. 1, § 25; MD. CONST. art. 47; MICH. CONST. art. 1, § 24; MISS. CONST. art. 3, § 26A(1); MO. CONST. art. 1, § 32; NEB. CONST. art. I, § 28; NEV. CONST. art. 1, § 8; N.J. CONST. art. 1, ¶. 22; N.M. CONST. art. II, § 24; N.C. CONST. art. I, § 37; OHIO CONST. art. I, § 10a; OKLA. CONST. art. 2, § 34; OR. CONST. art. I, § 42; R.I. CONST. art. 1, § 23; S.C. CONST. art. I, § 24; TENN. CONST. art. I, 35; TEX. CONST. art. 1, § 30; UTAH CONST. art. I, § 28; VA. CONST. art. I, § 8-A; WASH. CONST. art. I, § 35; Wis. Const. art. 1, § 9m.

218. For a more in depth history of Congress's passage of victims' rights laws and the collateral attempt to pass a victims' rights amendment to the U.S. Constitution, *see* Aaronson, *supra* note 211, at 626–34; Cassel, *supra* note 211, at 865–70; Kyl et al., *supra* note 3, at 583–91.

219. 150 CONG. REC. S10910 (daily ed. Oct. 9, 2004) (statement of Sen. Kyl) (arguing that "[v]ictims are the persons who are directly harmed by the crime and they have a stake in the criminal process because of that harm.").

220. *See* Beloof, *Weighing Crime Victims' Interests*, *supra* note 158, at 1149 ("The

York was not kidnapped, beaten, and raped. I was.'"[221] Second, advocates have argued that the criminal justice system should seek ways to avoid imposing secondary harm on victims during the prosecutorial process.[222] It is here that Professor Douglas E. Beloof, a leading victims' rights advocate and scholar, has suggested that, in conjunction with Professor Packer's models for the criminal justice system,[223] a third model should be added: the Victim Participation Model.[224]

In contrast to the Crime Control Model, which values the efficient prosecution of crime for the benefit of society, and the Due Process Model, which values the procedurally fair prosecution of crime for the benefit of defendants, the core value of the Victim Participation Model is victim primacy.[225] Inherent in recognizing the primacy of the victim is an acknowledgment that the victim should be treated with dignity, respect, and fairness.[226] In this regard, the Victim Participation Model asserts that victims should be granted "due-process like" rights to actively participate in the prosecution of the offender.[227] By providing victims participatory rights, the Victim Participation Model seeks to alleviate the two core harms suffered by victims:

> The first harm is primary harm, which results from the crime itself. The other harm is secondary harm, which comes from governmental processes and governmental actors within those processes. . . . The primary harm is a basis for victim participation in the same way that harm to an individual, coupled with a legitimate theory of the liability of another, is the basis for standing in other legal contexts. The potential for secondary harm

---

federal and state victims' rights laws legitimize crime victim harm upon which victims' interests in justice and minimizing secondary victimization are based."); Deborah P. Kelly, *Victims' Perceptions of Criminal Justice*, 11 PEPP. L. REV. 15, 18–20 (1984) (noting that victims want their personal interests to be recognized by the system and to be more involved in the process); Robert P. Mosteller, *Victims' Rights and the Constitution: Moving from Guaranteeing Participatory Rights to Benefiting the Prosecution*, 29 ST. MARY'S L.J. 1053, 1054 (1998) (noting one justification for victims' rights laws is based on granting to victims more participatory rights in the process).

221. Cellini, *supra* note 16, at 850 (quoting Paul S. Hudson, *The Crime Victim and the Criminal Justice System: Time for a Change*, 11 PEPP. L. REV. 23, 25–26, n. 14 (1984)).

222. *See supra* notes 215–216 and accompanying text; *see also* Beloof, *The Third Model*, *supra* note 175, at 294–96 (noting that the criminal justice system should take into consideration minimizing secondary harm to the victim); Beloof, *Weighing Crime Victims' Interests*, *supra* note 158, at 1149 (noting that a goal of victims' rights laws is to minimize secondary victimization); Kelly, *supra* note 220, at 21 (advocating for increased victim participatory rights in the criminal justice process).

223. *See supra* notes 174–193 and accompanying text.

224. Beloof, *The Third Model*, *supra* note 175, at 292.

225. *Id.* at 295–96.

226. *Id.* at 293.

227. *Id.* at 294. These include the "rights to notice and attendance, and the right to speak to the prosecutor and the judge." *Id.*

> provides a significant basis for a victim's civil rights against governmental authority.[228]

Victim participation in the prosecutorial process acknowledges both of these harms, and in so doing, honors the victim.

The question remains how to effectively integrate the Victim Participation Model into a public prosecution system, where utilitarian and retributive foundations have largely disregarded the victim. It is here where I believe procedural justice can serve as the appropriate conduit to expand both the utilitarian and retributivist approaches to the public prosecution model and to provide the victim, through Professor Beloof's Victim Participation Model, a more grounded role and place within the criminal justice system.

D. Procedural Justice and the Victim

Procedural justice theory generally posits that an individual's evaluation of the fairness of a decision is not based only on the final conclusion reached by decision makers, but also on the process by which the authorities reached that conclusion.[229] This discipline evolved out of social science research that sought to evaluate litigant satisfaction in case outcomes based on the process by which those outcomes were reached.[230] As procedural justice theory has evolved, its focus has broadened to include concerns not only about a fair decision-making process, but also concerns

---

228. *Id.* at 294–95 (citations omitted).

229. *See infra* notes 233–247 and accompanying text.

230. Jonathan D. Casper, Tom Tyler & Bonnie Fisher, *Procedural Justice in Felony Cases*, 22 LAW & SOC'Y REV. 483, 486–87 (1988); *see also* Rebecca Hollander-Blumoff & Tom R. Tyler, *Procedural Justice in Negotiation: Procedural Fairness, Outcome Acceptance, and Integrative Potential*, 33 LAW & SOC. INQUIRY 473, 477 (2008) (outlining basic principles of procedural justice); Elizabeth Chamblee Burch, *Procedural Justice in Nonclass Aggregation*, 44 WAKE FOREST L. REV. 1, 26–46 (2009) (discussing fundamental procedural justice components).

It should be noted the concept of procedural justice should not be used as shorthand for procedural due process, or as a substitute for the established rules of criminal procedure. Rather, procedural justice is a broader concept that seeks to evaluate individual perceptions of the fairness regarding official decision making processes and their associated outcomes. Of course, it is generally accepted that if a criminal proceeding is held in accordance with the Rules of Criminal Procedure, and affords the defendant procedural due process, the proceeding and its end result can be deemed to be fair. However, because the law is unwilling to formally extend to victims a protectable legal interest in criminal proceedings in the same fashion as is granted to the state and defendant, invocations of procedural due process and standard legal procedure are misplaced. However, under the broader umbrella of procedural justice, crime victims can nonetheless make a case for exercising more of a voice in criminal actions.

about whether those impacted by the decision were treated fairly throughout the process.[231]

In their earliest iterations, procedural justice theorists focused on the intersection between an individual's evaluation of an official decision and the process by which that decision was reached.[232] As advanced by John Thibaut and Laurens Walker, this "process control" approach to procedural justice asserts that

> [a] process in which litigants feel that they have the opportunity to express their point of view fully and in which the decision maker is perceived as having listened to and considered their side's arguments will promote a sense of fair treatment and thus a sense of satisfaction with the court experience.[233]

Studies indicated that individuals who received unfavorable outcomes, but who nonetheless perceived that they were able to fully express their views during the proceeding, were more satisfied with their overall experience than those who received favorable outcomes, but had less of an opportunity to express their views.[234] At a minimum, the former were more likely to think the outcome was fair because of their level of participation in the process.[235]

But why should it matter that a process is perceived to be fair? Procedural justice theorists contend that a fair process helps the participant shape his or her beliefs about the legitimacy of those making the decisions

---

231. *See infra* notes 248–275 and accompanying text.

232. *See, e.g.*, Casper et al., *supra* note 230, at 485–87 (discussing "[a] variety of factors [that] are typically said to influence a citizen's satisfaction with an encounter with a legal institution, including case outcome, distributive justice, and procedural justice.").

233. *Id.* at 486.

234. *Id.* at 486–87; *see also* Tom R. Tyler, *Does the American Public Accept the Rule of Law? The Findings of Psychological Research on Deference to Authority*, 56 DEPAUL L. REV. 661, 663 (2007) [hereinafter *The Findings of Psychological Research*]. Six criteria are relevant in evaluating the fairness of a legal proceeding. Tom R. Tyler, *What is Procedural Justice: Criteria Used by Citizens to Assess the Fairness of Legal Procedures*, 22 LAW & SOC'Y REV. 103, 104–05 (1988) [hereinafter *What is Procedural Justice*]. They include consistency, bias suppression, accuracy of information, the ability to appeal or correct wrong decisions, the ability to have access or input at all levels of the decision making process (sometimes termed "representativeness"), and ethicality. *Id.* The last factor, ethicality, "refers to the degree to which the decision-making process accords with general standards of fairness and morality." *Id.* at 105. The six factors were identified by Gerald S. Leventhal, who built upon the work of John Thibaut and Laurens Walker. *See* Hea Jin Koh, *"Yet I Shall Temper So Justice with Mercy": Procedural Justice in Mediation and Litigation*, 28 LAW & PSYCHOL. REV. 169, 170 (2004); Tyler, *What is Procedural Justice*, *supra* note 234, at 104–05.

235. *See* Tyler, *The Findings of Psychological Research*, *supra* note 234, at 664 (suggesting that "hav[ing] an opportunity to state their case" makes a legal process fair in the eyes of the public).

and subsequently leads to the participant's increased compliance with the law and future cooperation with authorities.[236] For example, in a 1997 study, researchers examined the impact fair procedures had on spousal assault recidivism rates.[237] The study revisited surveys completed in Minneapolis in the early 1980s, which suggested that if police arrested individuals charged with spousal abuse, rather than taking other actions such as mandating a "walk around the block" to ease tensions or issuing a warning to the parties, the arrestees would engage in fewer subsequent violent acts.[238] The 1997 follow-up study found that when a defendant was arrested, but felt like the process he or she received was fair, the arrestee's rate of recidivism was lower than for those defendants who did not believe they encountered fair procedures.[239] Similarly, the rate of recidivism between those who were arrested but perceived that they were treated fairly, and those who were not arrested, was similar.[240] Therefore, fair procedures, even if associated with unfavorable outcomes, were more likely to result in long-term compliance with the law.[241] Hence, from a utilitarian perspective, a system that is perceived as fair may ultimately result in a greater social good. The citizenry will have more faith in the system, be more willing to cooperate with the police and follow the law, and be more likely to comply with any individual sanctions imposed on them.[242]

One might question, however, whether these process control benefits matter to victims. This strain of procedural justice seeks to gain (or maintain) the favor of those who have been sanctioned in some manner. The hope is that by treating these individuals fairly, the utilitarian goals of specific and general deterrence are furthered.[243] A fair process and punishment are more likely than a seemingly unfair process or punishment

---

236. *See id.* at 676–77.

237. Raymond Paternoster, Robert Brame, Ronet Bachman & Lawrence W. Sherman, *Do Fair Procedures Matter? The Effect of Procedural Justice on Spouse Assault*, 31 LAW & SOC'Y REV. 163, 163 (1997).

238. *Id.* at 163–64.

239. *Id.* at 190.

240. *Id.* at 163.

241. The inverse, therefore, would also be true. "People who have experienced a procedure they judge to be unfair are not only less respectful of the law and legal authorities, they are less likely to accept judicial decisions and less likely to obey the law in the future." Tom R. Tyler, *The Psychological Consequences of Judicial Procedures: Implications for Civil Commitment Hearings*, 46 SMU L. REV. 433, 439 (1992) [hereinafter *Psychological Consequences*].

242. *See* Tom R. Tyler & Jeffrey Fagan, *Legitimacy and Cooperation: Why do People Help the Police Fight Crime in their Communities?*, 6 OHIO ST. J. CRIM. L. 231, 233 (2008) (noting that citizens' views of the legitimacy of law enforcement shapes the extent of their cooperation with police).

243. *See, e.g.*, Paternoster et al., *supra* note 237, at 193–94 (discussing theories that suggest that "compliance is more likely when authorities impose sanctions while still honoring and respecting the dignity of offenders.").

to result in future compliance and cooperation with the law. However, process control appears to be misplaced when applied to victims. It is not necessarily the victim who needs to be deterred, whether specifically or generally, from committing crime.[244] Therefore, when questioning the value of the process control model of procedural justice to victims, one must reframe the argument.

Even though the public prosecution model has diminished the victim's place within the criminal justice process, victims still play an important role in an efficiently functioning system. For example, it is often the victim who reports the crime, provides relevant information to the police to assist them in their efforts to apprehend the perpetrator, and serves as a witness for prosecutors.[245] Victim experiences that do not comport with procedural justice principles may lead victims to question the system's legitimacy and undermine their willingness to cooperate with authorities in the future.[246] Hence, pursuant to the purely utilitarian goal of crafting a criminal justice

---

244. Professor Michael M. O'Hear articulates a valid challenge to the common assumption that victims are wholly blameless and free from any personal taint of crime. Michael M. O'Hear, *Plea Bargaining and Victims: From Consultation to Guidelines*, 91 MARQ. L. REV. 323, 327 (2007). He notes that

> many victims are themselves involved in criminal activity, live in neighborhoods with high crime rates, or are otherwise at high risk for involvement in or exposure to additional offenses. As to such victims in particular, there may be important law enforcement benefits that result from perceptions that legal authorities are worthy of respect and cooperation. (One need only think of the "anti-snitching" movement found in many minority communities to appreciate the consequences of a breakdown in respect for the authorities.)

Id. at 327–28 (citation omitted). Hence, there may be a positive deterrent outcome from treating victims fairly.

245. Kelly, *supra* note 220, at 20–21 (noting a study indicating that 87% of reported crime comes to police attention through victim reports); *see also* Beloof, *The Third Model*, *supra* note 175, at 306 (citing statistics indicating that many victims decide not to report crime); McDonald, *supra* note 11, at 650 (stating that "[t]he only concern that the state has for the victim is his willingness to cooperate and his ability to be a convincing witness"). The information provided by victims can be vital to any number of aspects of the criminal process, including framing plea bargains and sentencing decisions. *See generally* O'Hear, *supra* note 244 (discussing the benefits which may arise from providing victims with procedural justice during the plea process). Moreover, victim input at sentencing has the capacity of ensuring that the defendant receives a just sentence, therefore furthering retributive goals of the criminal justice system. *See supra* notes 194–200 and accompanying text.

246. *See* Cellini, *supra* note 16, at 851 (citing studies that have shown "negative encounters with the [criminal] justice system cause victims to opt out of future cooperation.").

system that efficiently responds to criminal wrongdoing, a system that provides victims with procedural justice is important.

Moreover, a criminal justice system that successfully harnesses the individual and collective desires of the citizenry to "bring the offender to justice" will maintain public confidence. Recall Professor Packer's statement regarding the Crime Control Model: "[t]he failure of law enforcement to bring criminal conduct under tight control is viewed as leading to the breakdown of public order and thence to the disappearance of an important condition of human freedom."[247] Simply put, from a utilitarian viewpoint effective prosecutions can result in happy—or at least satisfied—victims. Satisfied victims, in turn, help comprise a satisfied citizenry. A satisfied citizenry will continue to support and comply with the rules and processes of their self-created governmental structures.

Procedural justice theory, however, has not focused solely on the question of whether individuals perceive the outcome of their case to be fair based on the processes that led to that outcome. The discipline has also evolved to acknowledge that the process, in and of itself, has great value to those who participate in it.[248] This second layer of procedural justice theory asserts that fair treatment should not merely be viewed as a means to an end, such as reaching a favorable outcome, but as an end in itself.[249] This alternative approach to procedural justice has been described by Professors Tom Tyler and Allan Lind as the "group value theory."[250] The group value theory asserts that "people care whether their treatment (and not simply their outcomes) is fair because fair treatment indicates something critically important to them—their status within their social group."[251]

The core premise of the group value theory is that "people are predisposed to belong to social groups and that they are very attentive to signs and symbols that communicate information about their position

---

247. Packer, *supra* note 174, at 9.

248. Larry Heuer, Eva Blumenthal, Amber Douglas & Tara Weinblatt, *A Deservingness Approach to Respect as Relationally Based Fairness Judgment*, 25 PERSONALITY & SOC PSYCHOL. BULL. 1279, 1280 (1999) (discussing evolution of procedural justice from a process control model to a group value theory); Adam Lamparello, *Social Psychology, Legitimacy, and the Ethical Foundations of Judgment: Importing the Procedural Justice Model to Federal Sentencing Jurisprudence*, 38 COLUM. HUM. RTS. L. REV. 115, 124–126 (2006) (discussing the "empirical shift in focus from distribution to procedural justice"); Dale T. Miller, *Disrespect and the Experience of Injustice*, 52 ANN. REV. PSYCHOL. 527 529–30 (2001) (discussing shift in focus of procedural justice research).

249. *See* Lamparello, *Social Psychology supra* note 248, at 128–29; Miller, *supra* note 248, at 529.

250. The group value model is also sometimes referred to as the relational model of procedural justice. *See* Heuer et al., *A Deservingness Approach*, *supra* note 248, at 1280; Tom R. Tyler, *Psychological Models of the Justice Motive: Antecedents of Distributive and Procedural Justice*, 67 J. PERSONALITY & SOC. PSYCHOL. 850, 851–52 (1994) [hereinafter *Psychological Models*].

251. Miller, *supra* note 248, at 529.

within groups. . . . People want to understand, establish, and maintain social bonds."[252] Being part of a group provides a source of self-validation and further evidence that one is a fully accepted member of their group, which is rewarding, just as it is troubling when individuals perceive that they are being rejected or excluded.[253] This becomes all the more true when one perceives having been accepted or rejected by an authority figure, such as the state.[254] In this regard, individuals

> respond to whether they are treated with respect, politeness, and dignity, and whether their rights as citizens are acknowledged. People value the affirmation of their status by legal authorities as competent, equal, citizens and human beings, and they regard procedures as unfair if they are not consistent with that affirmation. . . . [I]t is important to recognize that government has an important role in defining people's views about their value in society.[255]

In examining how an individual evaluates the treatment he or she receives from authorities, procedural justice theorists have proffered a variety of formulations[256] that can be summarized into four core categories. First, it is important that individuals are provided with an opportunity to tell their side of the story or use their own voice.[257] Second, those overseeing the judicial process should be neutral.[258] In this regard, one should ask

---

252. Tyler, *Psychological Models*, *supra note* 250, at 851; *see also* Burch, *supra* note 230, at 19–20 (2009); Deborah Epstein, *Procedural Justice: Tempering the State's Response to Domestic Violence*, 43 WM. & MARY L. REV. 1843, 1875–77 (2002), Adam Lamparello, *Incorporating the Procedural Justice Model into Federal Sentencing Jurisprudence in the Aftermath of* United States v. Booker: *Establishing United States Sentencing Courts*, 4 N.Y.U. J.L. & LIBERTY 112, 123 (2009); Heather Smith et al., *The Self-Relevant Implications of the Group-Value Model: Group Membership, Self-Worth, and Treatment Quality*, 34 J. EXPERIMENTAL SOC. PSYCHOL. 470, 471 (1998) (discussing the group value theory of procedural justice).

253. Tyler, *Psychological Models*, *supra* note 250 at 851–52.

254. *See id.* at 852.

255. Tyler, *Psychological Consequences*, *supra* note 241, at 440–41 (citations omitted); *see also* Larry Heuer, *What's Just About the Criminal Justice System? A Psychological Perspective*, XIII J.L. & POL'Y 209, 211 (2005) (explaining that the group value model "assumes that group identification is psychologically rewarding and that individuals are motivated to establish and maintain group bonds."); Paternoster et al., *supra* note 237, at 167 (suggesting that "persons who are treated fairly feel attached to the social order, . . . they perceive that they are valued members of the group.").

256. *See, e.g.*, Burch, *supra* note 230, at 28–29 (referencing six Leventhal factors often referenced in process outcome settings); Epstein, *supra* note 252, at 1876–77 (detailing a four factor approach and including the elements of voice, neutrality, consistency, and dignity); Tyler, *Psychological Models*, *supra* note 250, at 853 (describing three core factors: neutrality, trust, and standing).

257. Epstein, *supra* note 252, at 1876, 1878; Heuer, *supra* note 255, at 211.

258. Burch, *supra* note 230, at 28; Epstein, *supra* note 252, at 1877.

whether the official creates an even playing field among the participants by treating everyone fairly.[259] Therefore, "[n]eutrality involves honesty and a lack of bias. Neutral decision making also uses facts, not opinions, leading to decisions of objectively high quality."[260] Third, authorities should appear to be caring and trustworthy.[261] This factor serves to bolster participant confidence in the consistency and fairness of the decision makers: "If people are able to infer a benevolent disposition [in the decision maker], they can trust that in the long run the authority with whom they are dealing will work to serve their interests."[262] Finally, individuals value being treated with dignity, politeness, and respect.[263] Taken together, these factors highlight that there is something symbolically important about providing an individual with the opportunity to state her case, to have her statements received with an open mind and be taken seriously by decision makers. Such treatment not only contributes to an individual's sense of self-worth and standing within her social group, but also enhances her perceptions that the authorities making the decisions are moral and legitimate.[264]

The group value approach to procedural justice gets to the heart of many of the asserted goals of victims' rights laws. A process that transmits a message that the victim has worth and standing within the social group helps not only to temper secondary victimization but also to neutralize the negative social message broadcast by the defendant about the victim's value when the crime was committed.[265]

The group value variation of procedural justice can easily be incorporated into a utilitarian approach to criminal prosecutions. If crime represents, in some part, the offender's devaluing and marginalization of the victim, then a worthy goal of the criminal justice system should be to reverse that marginalization through the prosecution and punishment of the offender.[266] A legal system that treats victims with dignity and allows them to participate appropriately in the prosecution of the offender honors the individual victim's suffering while responding to the broader societal injury that resulted from the offender's crime.[267] Again, from a utilitarian perspective, the legal system's official acknowledgment of the victim's harm increases the overall social welfare of the society.[268] Additionally,

---

259. Tyler, *Psychological Models*, *supra* note 250, at 854.

260. *Id.*

261. Burch, *supra* note 230, at 28; Tyler, *Psychological Models*, *supra* note 250, at 854.

262. Tyler, *Psychological Models*, *supra* note 250, at 854 (citation omitted).

263. Burch, *supra* note 230, at 28; Epstein, *supra* note 252, at 1877.

264. *See* Epstein, *supra* note 252, at 1876; Heuer, *supra* note 255, at 211.

265. *See* George P. Fletcher, *What is Punishment Imposed For?*, 5 J. CONTEMP. LEGAL ISSUES 101, 110 (1994).

266. *See* Eisenstat, *supra* note 164 at 1154.

267. *See* Jamie Malamud Goti, *Equality, Punishment, and Self-Respect*, 5 BUFF. CRIM. L. REV. 497, 504 (2002).

268. *See* Sendor, *supra* note 193.

victims may feel less traumatized by the crime, suffer less secondary victimization, and therefore be more likely to maintain their trust and confidence in the legal system.[269] However, the group value approach to procedural justice can also accommodate retributive theories regarding the criminal process.

Driven largely out of a concern that official responses to crime should not be influenced by individual desires of revenge or vengeance, retributivists have tended to reject claims that the criminal justice system should acknowledge victims or their interests.[270] However, as retributivists move beyond their core premise that the society should punish only when and to the extent punishment is deserved and onto the larger precursor question of *why* such punishment is deserved, the victim inevitably becomes a part of the conversation.[271] Pursuant to social contract theory, we are collectively bound to certain social norms.[272] An individual's breach of those commonly accepted norms harms us all.[273] Therefore, the state is vested with the power to hold accountable the individuals who violate those collective social norms.[274] However, while the norms and the resulting harms may be viewed in the collective, the specific harm still tends to be directly predicated upon the particular injury suffered by an individual victim. Most criminal statutes could not exist without some reference to the harm suffered by an individual human being.[275] A specific woman was raped; a man was beaten; an investor was defrauded. The question thus becomes how retributive theory goes about acknowledging the victim.

Professors George P. Fletcher and Jean Hampton have both articulated forms of retributivism that seek to acknowledge the victim's place in the criminal justice equation. Although their arguments are distinct, they share the common feature that a retributivist response to crime should be motivated out of a desire to acknowledge the moral harm suffered by the victim and the correlative desire to correct that moral imbalance.

Professor Fletcher suggests that when an individual commits a crime, a particular relationship is established between the offender and the victim in

---

269. *See* Heuer, *supra* note 255.

270. *See* Christopher, *supra* note 180, at 937–38; *supra* notes 194–200 and accompanying text.

271. *See, e.g.*, Christopher, *supra* note 180, at 936–37; George P. Fletcher, *The Place of Victims in the Theory of Retribution*, 3 BUFF. CRIM. L. REV. 51, 55 (1999) [hereinafter *The Place of Victims*]; Goti *supra* note 267, at 504–08; Jean Hampton, *Correcting Harms Versus Righting Wrongs: The Goal of Retribution*, 39 UCLA L. REV. 1659, 1685–98 (1992); Moore *supra* note 193, at 69–70.

272. *See* Hampton, *supra* note 271, at 1694.

273. *Id.*

274. *Id.*

275. Christopher, *supra* note 180 at 935.

which "[t]he criminal gains a form of dominance [over the victim] that continues after the crime has supposedly occurred."[276] For example,

> [r]ape victims have good reason to fear that the rapist will return, particularly if the rape occurred at home or he otherwise knows her address. Burglars and robbers pose the same threat. Becoming a victim of violence beyond the law means that what we all fear becomes a personal reality; exposure and vulnerability take hold and continue until the offender is apprehended.[277]

The reason, then, that we arrest, try, and punish offenders, "is to overcome this dominance and reestablish the equality of the victim and offender."[278] A failure to respond to criminal wrongdoing would represent "abandoning victims in their suffering and isolation."[279] Our public prosecution system of criminal justice, therefore, should serve as a means for the entire community to stand in solidarity with the victim.[280] Here, Professor Fletcher invokes Kant's example of an island society about to disband, but still has offenders who have yet been brought to justice for their crimes. As described by Kant:

> Even if a civil society were to dissolve itself by common agreement of all its members (for example, if the people inhabiting an island decided to separate and disperse themselves throughout the world), the last murderer remaining in prison must first be executed, so that everyone will duly receive what his actions are worth and so that the bloodguilt thereof will not be fixed on the people because they failed to insist on carrying out the punishment; for if they fail to do so, they may be regarded as accomplices in this public violation of legal justice.[281]

Should the state fail to prosecute and punish the wrongdoers, it becomes complicit in continuing the defendant's asserted dominance over the victim and creates a new, and equally detrimental relationship of dominance and subservience between the victim and the state.[282] Thus, Professor Fletcher's

---

276. Fletcher, *The Place of Victims*, *supra* note 271, at 57; *see also* Fletcher, *What is Punishment Imposed For?*, *supra* note 265, at 109–10; Sendor, *supra* note 193, at 334, 354 (discussing Professor Fletcher's approach to retributivism).

277. Fletcher, *What is Punishment Imposed For?*, *supra* note 265, at 110; *see also* Fletcher, *The Place of Victims*, *supra* note 271, at 57–58 (illustrating the offender's dominance over the victim).

278. Fletcher, *The Place of Victims*, *supra* note 271, at 58.

279. Fletcher, *What is Punishment Imposed For?*, *supra* note 265, at 109.

280. *Id.*

281. *See* Immanuel Kant, THE METAPHYSICS OF MORALS 102 (John Ladd trans., 1965).

282. Fletcher, *The Place of Victims*, *supra* note 271, at 61–63; Fletcher, *What is Punishment Imposed For?*, *supra* note 265, at 110. In this regard, Professor Fletcher's dominance theory of retributivism shares many qualities with Professor Benjamin Sendor's articulation of restorative retributivism. *See* Sendor, *supra* note 193. Professor Sendor

dominance theory of retribution intersects nicely with a group value approach to procedural justice. By responding to criminal wrongdoing through public prosecutions and treating victims with dignity and respect throughout that process, the criminal justice system stands in solidarity with the victim and reaffirms human equality which was otherwise undermined by the defendant's wrongful acts.[283]

Professor Hampton's approach to retributivism echoes some of the themes appearing in Professor Fletcher's dominance theory, but Professor Hampton focuses more on the expressive nature of retributive responses to crime.[284] According to Professor Hampton, a defendant's criminal acts send a message, not only to the victim, but to the rest of society, that the victim possesses less value than she should otherwise be accorded.[285] By holding the wrongdoer accountable for his actions, we should seek to "vindicate the value of the victim denied by the wrongdoer's action through the construction of an event [like a trial] that not only repudiates the acto[r]'s message of superiority over the victim but does so in a way that confirms

---

contends that when the criminal justice system responds to wrongdoing, its goal should include

> the restoration of a moral balance among the offender, the victim, and the community. In other words, under the restorative theory, punishment focuses on the nature of the crime as an injury to relationships among the offender, the victim, and the community and, consequentially, it sends messages about right and wrong conduct to the victim and the community as well as to the offender.

Id. at 351.

It must be noted, however, that Professor Fletcher's approach to retributivism still does not fully integrate the individual victim into the criminal justice process. When discussing the victim, he does not focus on the needs and interests of a particular victim of a specific crime. Instead, he contends that "[t]he purpose of bringing victims into the analysis is not to hear their particular grievance and sentiments toward the offender, but to simply recognize that crime is first and foremost an action that causes harm to other people." Fletcher, *The Place of Victims*, *supra* note 271, at 55. Rather, what matters to him is "the victim-type, the victims of a class of those who have suffered a particular crime." *Id.* Therefore, the victim should be included in the criminal process not to vindicate that specific individual's harms, but to serve as representative of the general class of individuals who have suffered a criminal invasion of their interests. *Id.*

283. *See* Fletcher, *What is Punishment Imposed for?*, *supra* note 265, at 110; Heuer, *supra* note 255, at 211.

284. *See* Hampton, *supra* note 271; *see also* Laura I. Appleman, *Retributive Justice and Hidden Sentencing*, 68 OHIO ST. L.J. 1307, 1335–36 (2007) (explaining Hampton's "expressive" theory of retribution); Guyora Binder, *Victims and the Significance of Causing Harm*, 28 PACE L. REV. 713, 735 (2008) (describing Hampton's interpretation of retribution); Sendor, *supra* note 193, at 328, 335, 351–52 (summarizing Hampton's view on the goal of retribution).

285. *See* Hampton, *supra* note 271, at 1671.

them as equal by virtue of their humanity."[286] According to Professor Hampton, when reacting to criminal actions, "we are morally required to respond by trying to remake the world in a way that denies what the wrongdoer's events have attempted to establish, thereby lowering the wrongdoer, elevating the victim, and annulling the act of diminishment."[287] Hence, even if retributivism seeks to ensure that defendants are only punished for that which they justly deserve, Professor Hampton's expressive retributivism acknowledges that our individual and collective outrage at crime can, and should, be incorporated into how our criminal justice system responds to and punishes offenders.[288] It therefore remains vitally important that the state be involved in framing that reaction because how the state treats the victim is equally as potent as how the offender treated the victim. If criminal behavior is expressive, and hence sends message of degradation to those who observe it, "the state's behavior in the face of an act of attempted degradation against a victim is itself something that will either annul or contribute further to the diminishment of the victim."[289]

Both Professor Hampton's and Professor Fletcher's approaches to retributivism focus on affirming the value of the victim and therefore dovetail beautifully with the group value model of procedural justice and Professor Beloof's Victim Participation Model.[290] When victims are permitted to use their voice, encounter neutral and trustworthy decision makers, and are treated with dignity and respect during criminal proceedings, the goal of redeeming the denigration suffered by the victim can be fulfilled. These procedural justice practices can minimize any secondary harm suffered by victims, and simultaneously send a positive message to the victim, the defendant, and society regarding the victim's value.

The public prosecution model is firmly established as the manner by which we respond to crime.[291] Founded upon social contract theory and emanating from a desire to cabin victim vengeance, the public prosecution model has focused on finding the appropriate balance between society's desires to control crime and respond to criminal wrong doing and the individual defendant's liberty and due process interests. The result, however, was that the system drifted from its goal of tempering victim vengeance to ignoring the victim altogether. This need not be the case. By viewing utilitarian and retributivist criminal theory through the prism of

---

286. *Id.* at 1686.

287. *Id.* at 1686–87.

288. *Id.*

289. *Id.*

290. *See* Fletcher, *The Place of Victims*, *supra* note 271, at 58; Hampton, *supra* note 271, at 1686.

291. Juan Cardenas, *The Crime Victim in the Prosecutorial Process*, 9 HARV. J.L. & PUB. POL'Y 357, 371 (1986).

procedural justice principles, there is ample room to integrate the victim into the prosecutorial process.

IV. A REMEDY: PROCEDURAL JUSTICE AND THE PROMISE OF PROTECTION

Many of the rights afforded to victims under the CVRA already correspond nicely with procedural justice theory. Victims have the right to timely and accurate notice of public court proceedings, including any parole proceeding involving the crime or "any release or escape of the accused," the right not to be excluded from such public proceedings, the right to be reasonably heard at public proceedings regarding release, plea, sentencing or parole determinations, the "right to confer with the attorney for the Government," and "the right to proceedings free from unreasonable delay."[292] Should the prosecutor or court fail to provide a victim with these rights, the victim can file a writ of mandamus with the court of appeals, which must in a timely fashion grant or deny the writ.[293] If the victim's writ is denied, the court of appeals must issue a written opinion clearly stating the reasons for denial.[294] In allegiance with procedural justice principles, these rights, and the means by which victims can enforce them, ensure that victims can have a voice and role in the prosecution of the person alleged to have harmed them, wholly separate from any witness role the victim serves for the state.[295]

However, none of the rights in the statute give victims the power to control the proceedings or obtain a specific outcome. The CVRA makes clear that the failure to afford a victim the rights embodied in the statute does not establish grounds for a new trial against the defendant. [296] Rather, the rights afforded to victims give them an independent voice, rather than a veto over the criminal proceedings.[297] In this regard, the CVRA remains true to one of the core precepts of the public prosecution model by acknowledging that a criminal action represents a contest between the state and the defendant.[298]

By making clear that the victim merely possesses a voice, and not a veto in the criminal justice process, the CVRA appears to lean more towards embodying a group value rather than a process control approach to

---

292. 18 U.S.C. § 3771(a) (2006).

293. *Id.* at § 3771(d)(3).

294. *See id.* (explaining the statute's requirement to file and the court's obligation).

295. *See* Aaronson, *supra* note 211, at 662–66 (explaining further victims' right to have a role in the prosecution).

296. 18 U.S.C. § 3771(d)(5).

297. United States v. Turner, 367 F. Supp. 2d 319, 331 (E.D.N.Y. 2005) ("[T]he new law gives crime victims a voice but not a veto."); Kyl et al., *supra* note 3, at 622.

298. *See* Aaronson, *supra* note 211, at 675–78 (detailing the core precepts of the prosecution model and illustrating how the CVRA and victims fit into the picture).

victim procedural justice rights.[299] Of course, the utilitarian goals underlying the process control model are furthered if victims continue to cooperate with the criminal justice system despite the outcome of the case because they believe they were treated fairly.[300] However, I believe it is the dignitary interests embodied within the group value model that give the CVRA its greatest weight. The commonly asserted reasons for the necessity of victims' rights laws are that our criminal justice system needs to respond to individual victims' harms and to diminish any secondary harm that might befall the victims by their participation in the prosecution of the offender.[301] It is telling, therefore, that the last right listed for victims under the CVRA is "[t]he right to be treated with fairness and respect for the victim's dignity and privacy."[302] This language has been interpreted by some as vesting in victims specific and tangible rights,[303] but one could also read this language as signaling that when one views the CVRA rights as a whole, they serve to further a group value model approach to procedural justice. By giving victims independent participatory rights in the criminal justice system, the CVRA sends a powerful state-sanctioned message. The state broadcasts that victims are not merely a means to its goal of prosecuting the offender, but an end, in and of themselves, worthy of recognition and affirmation in the criminal justice process.[304]

We must question then, how the victim's right to reasonable protection from the accused can be brought within the fold of procedural justice and into alignment with the other CVRA rights. In contrast to the other rights in the CVRA, the statute's promise of protection has little to do with directly affirming the victim's dignity through participation in the criminal proceedings. Rather, the protection right suggests that the victim is a passive recipient of an assured government benefit.[305] Hence the right attempts to guarantee a specific outcome, such as protection from harm caused by the defendant, which can be incredibly difficult to ensure. Grounding the victim's right to reasonable protection from the accused within a procedural justice framework helps overcome these problems.[306]

---

299. Kyl et al., *supra* note 3, at 622.

300. *See id.* at 613–14 (explaining how victim participation can further the utilitarian goals and promote fairness).

301. *See supra* notes 214–216 and accompanying text.

302. 18 U.S.C. § 3771(a)(8).

303. *See* Cassell, *supra* note 211, at 872–77; *see also* United States v. Rubin, 558 F. Supp. 2d 411, 427–28 (E.D.N.Y. 2008); United States v. Patkar, No. 06-00259, 2008 WL 233062, at *5–6 (D. Haw. Jan. 28, 2008); United States v. Heaton, 458 F. Supp. 2d 1271, 1272–73 (D. Utah 2006).

304. *See* Cassell, *supra* note 211, at 876–77 (recognizing the importance of treating victims fairly).

305. *See* 150 CONG. REC. S10910 (daily ed. Oct. 9, 2004) (statement of Sen. Kyl) (showing framer's intent to provide victims with a right to protection).

306. However, as noted earlier, providing a victim with a secure waiting area during court proceedings might create an environment where the victim would be more likely to

First, the right needs to be re-written in such a way to emphasize the victim's participatory role in the criminal justice process, rather than promising a specific outcome. Hence, the CVRA's statutory language in section 3771(a)(1) should be amended to read: "A crime victim has . . . the right to have the victim's safety considered in determining the defendant's release from custody." Recall that in concert with the CVRA's current grant to victims that they be reasonably protected from the accused, the statute also grants victims the right to reasonable notice of any public court or parole proceeding involving the crime or any release or escape of the accused,[307] the right not to be excluded from any such proceedings,[308] the right to be reasonably heard at any such proceedings,[309] and the right to confer with the attorney for the government on the case.[310] These are all process-based rights, which, when exercised in the context of a court's release consideration hearing for a defendant, would give meaning to a right which seeks to prevent the defendant from causing the victim further harm. If a victim can confer with the government lawyer on the case, and is on notice of upcoming parole or release hearings, the victim can decide whether to exercise the right to be heard at those proceedings. If so, the victim can share safety concerns with the court, which the court can, in turn, consider in making its release decision.

Curiously enough, earlier statutory formulations regarding the victim's right to be reasonably protected from the accused were phrased in terms that focused far more on the victim's safety and role in the process of determining the defendant's release, than on promising a direct right to protection. Two previously proffered versions of the language which eventually appeared in the CVRA read as follows: victims have the right to have "the safety of the victim considered in determining a [defendant's] release from custody[,]"[311] and the victim has the "right to adjudicative

---

exercise the participatory rights under the statute. *See supra* notes 129–130 and accompanying text.

307. 18 U.S.C. § 3771(a)(2) (2006).

308. *Id.* at § 3771(a)(3). The statute does indicate, however, that the victim can be excluded if it determines that the testimony offered by the victim "would be materially altered if the victim heard other testimony at the proceeding. *Id.* The statute also states:

> In a case where the court finds that the number of crime victims makes it impracticable to accord all of the crime victims the rights described [in the statute], the court shall fashion a reasonable procedure to give effect to this chapter that does not unduly complicate or prolong the proceedings.

Id. at § 3771(d)(2).

309. *Id.* at § 3771(a)(4).

310. *Id.* at § 3771(a)(5).

311. S.J. Res. 65, 104th Cong. (1996); *see supra* note 218 regarding the history of the passage of the CVRA.

decisions that duly consider the victim's safety."[312] Many state laws parallel this type of language. For example, Alaska grants victims the right to protection through the imposition of appropriate bail or conditions of release by the court, as well as the right to be heard at any proceeding where the accused's release from custody is considered.[313] Similarly, Colorado and Florida grant victims the right to information regarding the steps they can take to protect themselves from harassment or harm from the offender,[314] and Indiana and Maryland indicate that victim safety should be considered in the process of determining whether to release the defendant from custody.[315] Therefore, altering the statutory language of the CVRA so that it focuses on considering the victim's safety, rather than making an outright promise of protection, is not unreasonable or unprecedented. Moreover, such an approach furthers the goal of enhancing the victim's appropriate participation in the criminal process. By framing the victim's protection right in terms of victim participation, rather than a specific outcome such as protection, the victim's ability to enforce the right is also more assured.

At least one CVRA case has implied that a victim's right to be heard in the context of a release determination is enforceable. In *United States v. Turner*,[316] the court noted that the Government failed to inform the victims of the defendant's detention hearing under the Bail Reform Act.[317] Acknowledging that there would be other bail-related proceedings prior to

312. S.J. Res. 35, 107th Cong. (2002).

313. *See* ALASKA CONST. art. I, § 24; ALASKA STAT. § 12.61.010 (2008).

314. COLO. REV. STAT. § 24-4.1-302.5(1)(m) (2010) ("[e]ach victim of crime shall have . . . [t]he right to be informed about what steps can be taken by a victim or witness in case there is any intimidation or harassment by a person accused or convicted of a crime against the victim, or any other person acting on behalf of the accused or convicted person"); FLA. STAT. ANN. § 960.001(1)(c) (West 2004) ("A victim or witness shall be furnished, as a matter of course, with information on steps that are available to law enforcement officers and state attorneys to protect victims and witnesses from intimidation.").

315. IND. CODE ANN. § 35-40-5-4 (LexisNexis 2010) ("A victim has the right to have the victims' safety considered in determining release from custody of a person accused of committing a crime against the victim."); MD. CODE ANN. CRIM. PROC. § 11-203 (LexisNexis 2008) ("[T]he court . . . shall consider . . . the safety of the alleged victim in setting conditions of . . . the pretrial release of a defendant . . . .").

316. 367 F. Supp. 2d 319 (E.D. N.Y. 2005) (mail fraud).

317. *Id.* at 320. It must be acknowledged that, as specifically related to the victim's right to be reasonably protected from the accused, the facts of *Turner* are unremarkable as the defendant in that case was charged with mail and financial fraud. *Id.* at 320. Fraud victims are perhaps not as likely to believe their personal safety is jeopardized, as a victim of violent crime might. Moreover, in deciding to detain the defendant, the court noted it did so not because of the danger the defendant presented to anyone, but rather because of the lack of assurance that the he would appear later at trial as required. *Id.* at 321.

the defendant's trial, and acknowledging its obligation to ensure that victims are afforded their rights,[318] the court

> directed the government to provide all alleged victims of the charged offense a written summary (if not a transcript) of the proceedings to date as well as notification of their rights under the statute with respect to future proceedings, including notice of the next scheduled proceeding and of their right to be heard with respect to [the defendant's] application for release.[319]

At the next hearing, the prosecution affirmed to the judge that each alleged victim had been informed of the proceeding, but "that none had elected to attend and be heard with respect to [the defendant's] application for release."[320]

While the facts of *Turner* do not provide a direct example of specific enforcement of the victim's right to be reasonably protected from the accused, the court's actions are nonetheless worth note. From a procedural justice standpoint, the court acknowledged that victim involvement in the process was important, even if, as implied from the opinion, victim involvement might not weigh heavily in the court's final determination.[321] By acknowledging the victim's right to notice, to be present, and to be heard at the defendant's bail release proceeding, even if it did not appear that victim safety was an issue, the court gave a measure of weight to the victim's right to be reasonably protected from the accused.[322]

In order, however, to make the victims' re-framed protection right enforceable, the language of the CVRA needs to be altered in another respect. Currently, when discussing how victims can enforce their rights, the statute indicates that the victims can only petition to re-open plea and sentencing hearings.[323] However, other portions of the statute indicate that the victim has the right to be heard at *release*, plea, sentencing, and *parole* hearings.[324] Hence, as the statute currently reads, the victim has the right to be heard at several distinct moments during the criminal process, but can only seek relief for the denial of the right to be heard in some of those proceedings. In order to make the protection right enforceable, victims need

---

318. 18 U.S.C. § 3771(b)(1) (2006) ("[T]he court shall ensure that a crime victim is afforded the rights described in" the statute.).

319. *Turner*, 367 F. Supp. 2d at 321.

320. *Id.*

321. *Id.* at 332 (noting that even if the court is required to give the victim an opportunity to be heard, it is not compelled to deny the defendant's release pending trial, if the court is assured that a conditional release will assure the defendant's later appearance at trial as well as the safety of others).

322. *See generally id.* at 331–34 (showing the court's acknowledgement of rights granted by the CVRA).

323. 18 U.S.C. § 3771(d)(5).

324. *Id.* at § 3771(a)(4).

to be able to seek a re-hearing in all relevant situations where they have the right but were denied the opportunity to be heard regarding their safety.[325] Therefore, section 3771(d)(5) of the statute should be amended to include a victim's right to be re-heard at release and parole hearings. The suggested change is as follows: "In no case shall a failure to afford a right under this chapter provide grounds for a new trial. A victim may make a motion to re-open a release hearing, plea, sentence, or parole hearing."

Reframing the victim's protection right in this manner does not represent a radical change from how the law already seeks to take victim safety issues into account. The Federal Bail Reform Act requires that a court consider whether the defendant poses a danger to any person or the community in the course of determining whether a defendant can be released pending trial.[326] The Bail Reform Act also requires that the court consider whether the defendant's release raises "a serious risk that such person will obstruct or attempt to obstruct justice, or threaten, injure, or intimidate, or attempt to threaten, injure, or intimidate, a prospective witness or juror."[327] In evaluating whether it believes such a risk exists, the court is meant to consider the nature and circumstances of the offense charged, the weight of the evidence against the accused, the history and characteristics of the person, and "the nature of the seriousness of the danger to any person or the community that would be posed by the person's release."[328] While the Bail Reform Act does not specifically mention victims, its broader language regarding the safety of the community, and prospective witnesses and jurors, could certainly include within its scope victim safety concerns.

---

325. Requiring such a re-hearing right would not run afoul of the Bail Reform Act, which indicates that

> the [release] hearing may be reopened, before or after a determination by the judicial officer, at any time before trial if the judicial officer finds that information exists that was not known . . . at the time of the hearing and that has a material bearing on the issue of whether there are conditions of release that will reasonably assure the appearance of [the defendant] as required and the safety of any other person and the community.

Id. at § 3142(f). *See also infra* notes 326–334 and accompanying text.

326. 18 U.S.C. § 3142(e) ("If, after a hearing . . . the judicial officer finds that no condition or combination of conditions will reasonably assure the appearance of the person as required and the safety of any other person and the community, such judicial officer shall order the detention of the person before trial."). The statute further guides that under a specific set of circumstances, a rebuttable presumption will arise that "no condition or combination of conditions will reasonably assure the safety or any other person and the community." *Id.*

327. *Id.* at § 3142(f)(2)(B).

328. *Id.* at § 3142(g)(4).

There is very little direct case law regarding application of safety provisions of the Bail Reform Act. However, there is at least one decision in which the trial court took into consideration the victim's safety when deciding not to release the defendant pending trial. In *United States v. Zuni*, the defendant was charged with abducting and sexually assaulting his estranged common law wife.[329] He already had a history of charges for physical assault against his common law wife and others, combined with a history of failing to comply with no-contact and restraining orders.[330] In evaluating whether to permit the defendant's release pursuant to the Bail Reform Act, the court determined that there was no condition or combination of conditions that it could fashion which would reasonably assure the victims' safety.[331] In light of the defendant's prior actions against his common law wife, coupled with his failure to comply with previously issued no-contact orders, the court stated:

> If [the defendant] violates his conditions—a prospect all too likely given his past—he poses a high degree of danger to [the victim], given the history of their relationship and their children's residence with [the victim.] Also, law enforcement may not be able to intercept [the defendant] or warn [the victim] before he [might reach her home]. Allowing [the defendant] out of detention now would pose too great of a risk to [the victim].[332]

The court in *Zuni* did not reference the victim's right under the CVRA to be "reasonably protected from the accused,"[333] nor did it give any indication that the victim or a representative for the victim was present at the defendant's bail release hearing.[334] The case nonetheless suggests that courts do factor victim safety into their release decision analyses.

Reconfiguring the victim's protection right so that it centers on victim participation in the decision making process regarding the defendant's pre-trial release is supported by current practice.[335] Whether under the terms of the Bail Reform Act, or other portions of the CVRA that grant the victim the right to be heard on issues regarding the defendant's release prior to trial or on parole, the case law recognizes that victim safety is an important factor that contributes to a well-functioning and fair criminal justice system.[336] Altering the statute's language so that it better reflects these

---

329. United States v. Zuni, No. 05-02569, 2006 WL 4062888, at *5 (D.N.M. 2006).

330. *Id.* at *3–4.

331. *Id.* at *6.

332. *Id.* at *7.

333. 18 U.S.C. § 3771(a)(1) (2006).

334. *Id.* at § 3771(a)(4) (crime victim has the "right to be reasonably heard at any public proceeding in the district court involving release, plea, sentencing, or any parole proceeding").

335. *See* OFFICE FOR VICTIMS OF CRIME, *supra* note 130, at 33–35.

336. *See* United States v. Turner, 367 F. Supp. 2d at 323–25 (reflecting the value the

realities will allow the victim's protection right to shift from being an empty promise to being an enforceable right. Of course, some might resist my retreat from the protection language currently present in the CVRA. However, as I have argued, that language lacks any enforceable substance. Conversely, the narrower statutory phrasing I have crafted embodies a procedural justice approach to victims in the law. This new language focuses on the process by which victims are heard regarding their safety and provides a direct manner by which victims can enforce that right.

V. CONCLUSION

Since its passage in 2004, the CVRA has opened the door of the criminal justice system to victims. The statute allows victims to participate in the prosecutorial process, thereby diminishing the secondary harm they may have previously suffered under the public prosecution model, while at the same time affirming their status as individuals with a direct interest in the proceeding. So doing, most of the rights afforded to victims under the CVRA can be viewed as an embodiment of procedural justice principles. The process rights granted to victims serve, in part, to enhance victim confidence in respect for the criminal justice system. Inversely, but perhaps more importantly, the process rights granted to victims signal the respect and value the system seeks to extend to them. It is time, therefore, to better align the CVRA's protection right with these procedural justice principles. The CVRA's promise of protection dishonors victims by granting them an empty and unenforceable right. Conversely, a promise that victims will be heard on issues regarding their safety fills the right with substance and meaning, thereby fulfilling the goals of the CVRA and the victims' rights movement.

---

court places on victim safety and well-being).

LESS THAN PICTURE PERFECT: THE LEGAL RELATIONSHIP BETWEEN PHOTOGRAPHERS' RIGHTS AND LAW ENFORCEMENT

MORGAN LEIGH MANNING*

I. INTRODUCTION 106
II. RECENT INCIDENTS INVOLVING THE VIOLATION OF PHOTOGRAPHY RIGHTS 108
A. *Incidents Involving Law Enforcement Officials* 108
B. *Incidents Involving Private Security Guards* 112
C. *Common Themes* 114
III. PHOTOGRAPHERS' RIGHTS—THE EXISTING LEGAL FRAMEWORK 115
A. *Public Places* 116
1. The Right to Photograph People in Public Places 116
2. The Right to Photograph Structures Visible from Public Places 118
B. *Privately Owned Places Open to the Public* 119
C. *Observations* 120
IV. REMEDIES FOR VIOLATIONS COMMITTED BY LAW ENFORCEMENT OFFICIALS IN PUBLIC PLACES 121
A. *Declaratory Relief* 121
B. *Injunctive Relief* 123
C. *U.S.C. § 1983 Action* 124
1. Establishing a Constitutional Violation 125
2. Qualified Immunity 131
3. Proving Damages 137
4. § 1983 Summary 139
D. Bivens *Action* 141
1. General Obstacles to *Bivens* Recovery 143
2. Obstacles to *Bivens* Plaintiffs within the Context of Photography Rights 143
3. *Bivens* Summary 145
V. VIOLATIONS COMMITTED BY SECURITY GUARDS IN PRIVATELY OWNED PLACES OPEN TO THE PUBLIC 146
A. *Private Security Guards Acting under Color of State Law* 147

* Associate in civil litigation with Slayton, Bain & Clary, Lawrenceville, Virginia. J.D., University of Tennessee College of Law, 2010; B.A., University of Virginia, 2007. The author would like to thank Professor Glenn Harlan Reynolds for his guidance, support, and insight during the writing of this article.

B. State Tort Claims Available against Private Security Guards not Considered to be Acting under Color of State law 148
VI. SUMMARY OF REMEDIES 149
VII. IMPLICATIONS OF AN INADEQUATE REMEDIAL STRUCTURE 151
VIII. WHAT'S WRONG WITH THIS PICTURE? 152
A. The Erosion of Photography Rights Will Diminish Public Awareness, Transparency, and Accountability 152
B. The Erosion of Photography Rights Reflects Insulated Policymaking Void of Accountability 154
C. The Erosion of Photography Rights Has not Been and Will not Be Accompanied by an Increase in National Security or Public Safety 155
IX. POSSIBLE SOLUTIONS 156
A. Know Your Rights 156
B. Publicize Violations 157
C. Attorneys: Attack the Statute 157
D. An Organized Front on Behalf of News Media 158
E. Observe the Efforts of Photographers Abroad 158
F. Expand the § 1983 Definition of "Acting under Color of State Law" to Include Private Security Guards 159
G. New Legislation under the Fourteenth Amendment 159
X. CONCLUSION 160

I. INTRODUCTION

Threats to national security and public safety, whether real or perceived, result in an atmosphere conducive to the abuse of civil liberties. History is littered with examples: The Alien and Sedition Acts of 1798, the suspension of habeas corpus during the Civil War, the Palmer Raids during World War I, and McCarthyism in the aftermath of World War II.[1] Unfortunately, the post-9/11 world represents no departure from this age-old trend. Evidence of post-9/11 tension between national security and civil liberties is seen in the heightened regulation of photography; scholars have labeled it the "War on Photography"—a conflict between law enforcement officials and photographers over the right to take pictures in public places.[2] In many cases, police officers and private security guards have invoked blanket notions of "national security" to prohibit the press and private photographers from taking pictures of structures that are in plain view of

1. *Civil Liberties in War Time*, BILL MOYERS JOURNAL, http://www.pbs.org/moyers/journal/07132007/civilliberties.html (last visited Jan. 25, 2011).

2. *See, e.g.*, Bruce Schneier, *The War on Photography*, SCHNEIER ON SECURITY (June 5, 2008, 6:44 AM), http://www.schneier.com/blog/archives/2008/06/the_war_on_phot.html.

the general public.[3] In other cases, law enforcement officials have used broadly worded criminal statutes such as "obstruction of justice" or "interfering with a police officer" to prohibit the press and private photographers from taking part in what is constitutionally protected behavior.[4] A simple Google search reveals countless incidents of overzealous law enforcement officials detaining or arresting photographers and, in many cases, confiscating their cameras and memory cards, despite the fact that these individuals were in *lawful* places, at *lawful* times, partaking in *lawful* activities.[5]

For at least two reasons, the argument that the heightened regulation of the right to take pictures in public places enhances national security or public safety is deeply flawed. First, the prevailing evidence indicates that the perpetrators of past terrorist attacks never photographed their targets.[6] Why would they need to, after all? The Internet and modern technology have made it possible to obtain pictures of most structures, especially ones located in urban areas, with the click of a mouse. For example, Google Earth provides images of almost any address in the country from a variety of distances and angles.

Second, even if terrorists did photograph their targets, it would be totally impractical to try to stop them. Bruce Schneier, an internationally known security technologist and author, notes:

> Billions of photographs are taken by honest people every year, 50 billion by amateurs alone in the US. And the national monuments you imagine terrorists taking photographs of are the same ones tourists like to take pictures of. If you see someone taking one of those photographs, the odds are infinitesimal that he's a terrorist.[7]

Questioning for the purpose of identifying potential terrorists persons taking pictures of the Empire State Building in New York City or the White House in Washington, D.C., makes less sense than trying to find a needle in a haystack, because, chances are, the needle does not exist.

Even more troubling is the fact that the misconduct on behalf of officials is not always motivated by a good-faith belief that their actions will promote public safety. In some instances, it appears to be motivated by distrust, even hostility, towards the press and private photographers.[8] One reason for this distrust and hostility could be awareness on behalf of officials that photographers (or anyone with a cell-phone camera for that

---

3. *See, e.g.*, incidents described *infra* notes 11–16, 40–49, 64–73.

4. *See, e.g.*, incidents described *infra* notes 11–16, 40–49, 64–73.

5. *See, e.g.*, incidents described *infra* notes 50–54, 82–86.

6. *See* Bruce Schneier, *The War on Photography*, SCHNEIER ON SECURITY (June 5, 2008, 6:44 AM), http://www.schneier.com/blog/archives/2008/06/the_war_on_phot.html.

7. *Id.*

8. *See, e.g.*, incidents described *infra* notes 11–16, 40–49, 64–73.

matter) can expose police misconduct. Because the press and private photographers serve as a check on official authority, police officers and security guards have an incentive to limit the power of individuals with cameras.

To make matters worse, for at least three reasons the victims of police misconduct in the area of photography rights will have a difficult time obtaining an adequate remedy at law. First, few effective remedies exist for compensating victims of photography rights violations. Second, courts have much discretion in awarding the remedies that do exist and often show deference to officials. Third, the nature of photography renders it difficult to prove damages to the degree of certainty the law requires.

This article examines the so-called War on Photography and the remedies available to those who have been unlawfully detained, arrested, or have had their property seized for taking pictures in public places or private places open to the public. It discusses recent incidents that highlight the growing infringement of photography rights and the magnitude of the harm that law enforcement officials have inflicted, paying particular attention to the themes these events have in common. It explores the existing legal framework surrounding photography rights and the federal and state remedies available to those whose rights have been violated. It examines the adequacy of each remedy including: (1) declaratory and injunctive relief, (2) § 1983[9] and *Bivens*[10] actions, and (3) state tort remedies. It discusses the obstacles associated with each remedy and the reasons why these obstacles are particularly hard to overcome in the context of photography. It then argues that most, if not all, of the remedies discussed are either inadequate or altogether impractical considering the costs of litigation. Lastly, this article will discuss the reasons why people should be concerned about the War on Photography and possible ways to reverse the erosion of photography rights.

II. RECENT INCIDENTS INVOLVING THE VIOLATION OF PHOTOGRAPHY RIGHTS

A. Incidents Involving Law Enforcement Officials

In August 2002, four police cars arrived on the scene at a river in Northwest Portland, Oregon, to investigate a report of suspicious activity involving people of Middle Eastern descent taking pictures near Portland bridges.[11] The suspects' names were Emily and Jenny, two high school students taking pictures for an art exhibit at the Portland Institute for Contemporary Art as part of a project called "Northwest Artists Respond to

9. 42 U.S.C. § 1983 (2006).

10. Bivens v. Six Unknown Named Agents, 403 U.S. 388 (1971).

11. Cliff Collins, *Defending the Right to Shoot*, 69 OR. ST. B. BULL. 42, 42 (2009).

Global Warming."[12] They were taking pictures of oil-storage tankers by the river when a security guard confronted them, questioned them, and later called the police.[13] The girls explained what they were doing, but the security guard responded that, since 9/11, they were not allowed to "do things like take pictures of bridges anymore."[14] The police came and, after investigating the incident, contacted the FBI, describing the girls as "two Middle Eastern-looking teenagers taking pictures near Portland bridges."[15] The officers threatened to confiscate the girls' film and told them that they would be placed on the FBI terrorism watch list.[16]

In July 2006, police in Philadelphia, Pennsylvania, arrested Neftaly Cruz, a Penn State University senior, and threatened to charge him with "conspiracy, impeding an investigation and obstruction of justice."[17] Cruz said that he heard a commotion and walked out of his back door to find the street lined with police cars.[18] He pulled out his camera phone and took a picture of the action.[19] Moments later, an officer came to Cruz's back gate.[20] Cruz stated that the officer "opened the gate and took me by my right hand."[21] The officer then "threw [Cruz] onto a police car, cuffed him and took him to jail."[22] Cruz said that the "police told him that he broke a new law that prohibits people from taking pictures of police with cell phones."[23] After about an hour in jail, the police "told him he was lucky because there was no supervisor on duty, so they released him."[24]

In July 2009, Gordon Haire, formerly both a law enforcement officer and newspaper reporter, was on campus at the University of Texas Medical Branch (UTMB) at Galveston, passing time before a doctor's appointment.[25] While sitting at a table, he snapped a picture of a university police officer walking in his direction.[26] The officer approached Haire and informed him that "it was against the law to photograph the Galveston

---

12. *Id.*
13. *Id.*
14. *Id.*
15. *Id.*
16. *Id.*
17. *Man Arrested For Taking Cellphone Photo of Police Activity*, INFORMATION LIBERATION (July 28, 2006, 11:40 PM), http://www.informationliberation.com/?id=13834.
18. *Id.*
19. *Id.*
20. *Id.*
21. *Id.*
22. *Id.*
23. *Id.*
24. *Id.*
25. Carlos Miller, *Texas Cops Prohibit Photography, Forbid the Filing of Complaints Against Them*, PHOTOGRAPHY IS NOT A CRIME (July 28, 2009, 1:13 AM), http://carlosmiller.com/2009/07/28/texas-cops-prohibit-photography-forbid-the-filing-of-complaints-against-them/.
26. *Id.*

National Laboratory" as doing so presented a "security threat."[27] Haire had not photographed the laboratory,[28] and questioned the officer as to whether it was really against the law to photograph the laboratory.[29] The officer requested Haire's identification, thereafter relaying through his collar microphone Haire's full name and date of birth "in a loud voice."[30] He then informed Haire that "it was illegal to even photograph the sidewalk" and left.[31] The next day, Haire walked into the UTMB Police Department with intentions of filing a complaint.[32] He was told he needed to "produce a photo ID," and when he could not, he was ushered out of the police station.[33]

In November 2006, Seattle police arrested and jailed amateur photographer Bogdan Mohora for snapping a few photographs of police officers arresting a man.[34] Two officers took his camera, wallet, and satchel, arrested him, and took him to a holding cell at the Seattle Police Department.[35] Mohora was never charged, and the police never wrote an incident report on the arrest.[36] He was released an hour later and was told that he could have been charged "with disturbing the peace, provoking a riot or endangering a police officer."[37] The American Civil Liberties Union (ACLU) intervened on his behalf and the city's claim department agreed to pay Mohora eight thousand dollars.[38] Both officers were "disciplined with written reprimands for a lack of professionalism and poor exercise of discretion."[39]

In June 2007, Indianapolis police questioned Walter Miller, a NASA employee, while Miller was touring the city and taking pictures.[40] Miller

---

27. *Id.*

28. *Id.*

29. *Id.*

30. *Id.*

31. *Id.*

32. *Id.*

33. *Id.*

34. *See* Christine Clarridge, *Man Jailed in Photo Incident Awarded $8,000*, SEATTLE TIMES, Nov. 9, 2007, http://seattletimes.nwsource.com/html/localnews/2004003761_photographer09m.html; *see also* Kathleen Davis, *The Crime of Photography: Rewarded!*, POP PHOTO FLASH DAILY BLOG (Nov. 19, 2007, 12:05 PM), http://flash.popphoto.com/blog/2007/11/the-crime-of-ph.html.

35. *See* Clarridge, *supra* note 34.

36. *See id.*

37. *Id.*

38. *See id.*

39. *Id.*

40. Carlos Miller, *Indianapolis Police Claim It Is Unlawful to Photograph Government Buildings*, PHOTOGRAPHY IS NOT A CRIME, (June 26, 2007, 10:53 PM), http://carlosmiller.com/2007/06/26/indianapolis-police-claim-it-is-unlawful-to-photograph-government-buildings/#more-151 (citing Sandra Chapman, *Visitor Didn't Feel Hoosier Hospitality*, WTHR EYEWITNESS NEWS, http://www.wthr.com/story/6698469/visitor-didnt-

was photographing an art exhibit that happened to be outside the Indianapolis City County Building.[41] Two police cars drove up to Miller, "[o]ne on the side of [him] and one behind [him] with their lights flashing."[42] The officers asked Miller what he was taking pictures of, and Miller replied, "[w]ell, the art exhibit."[43] One of the officers asked to see his camera, stating "I need to see it, for matters of homeland security."[44] The officer added, "You can't be taking pictures around here."[45] According to the Indianapolis Metropolitan Police Department (IMPD), "pictures of certain government facilities are off limits."[46] After questioning him, the officers allowed Miller to leave.[47] IMPD officials "say law enforcement is concerned about pictures of federal office buildings, military installations, major bridges and other infrastructure that could be considered a terrorists [sic] target."[48] If in doubt, IMPD officers say, "tourists should confine their photographs to marked tourists [sic] spots."[49]

In October 2009, in Anne Arundel County, Maryland, Antonio Amador "grabbed his camera to photograph a fatal accident that took place outside his home."[50] He was taking the photographs as part of an ongoing project aimed at highway safety and speed reduction.[51] Amador stated, "[s]uddenly I hear this screaming, like somebody really mad."[52] He continued, "I see this guy charging at me saying, 'delete those pictures now!'"[53] The officers threatened to arrest him if he did not delete his photos, and Amador complied.[54]

In February 2007, Miami multimedia journalist Carlos Miller was taking pictures of police officers for an article he was writing.[55] Officers demanded that Miller stop taking pictures, describing their actions as "a private matter."[56] Miller replied that it was a "public road."[57] The officers

---

feel-hoosier-hospitality?redirected=true (last visited Jan. 25, 2011)).

41. *See* Miller, *supra* note 40.

42. Chapman, *supra* note 40.

43. *Id.*

44. *Id.*

45. *Id.*

46. *Id.*

47. *See id.*

48. *Id.*

49. *Id.*

50. Carlos Miller, *Maryland Cops Force Photographer to Delete Photos*, PHOTOGRAPHY IS NOT A CRIME (Nov. 6, 2009, 4:41 AM), http://carlosmiller.com /2009/11/06/maryland-cops-force-photographer-to-delete-photos/.

51. *Id.*

52. *Id.*

53. *Id.*

54. *Id.*

55. Carlos Miller, *About the Blog*, PHOTOGRAPHY IS NOT A CRIME, http://carlosmiller.com/about/ (last visited Jan. 25, 2011).

56. *Id.*

then escorted Miller across the street and ordered him to keep walking away from the scene of the investigation.[58] Miller stated, "When I refused and continued to take their photo, they tackled me and bashed my head against the pavement, breaking a $400 camera flash and threatening to shoot me with a Taser gun."[59] Miller spent sixteen hours in the county jail on nine misdemeanor counts, the main charge being "obstructing traffic."[60] After sixteen months of delays, Miller finally went to trial.[61] After two days, Miller was acquitted of all criminal charges.[62] He was, however, found guilty of resisting arrest without violence, a charge he appealed pro se.[63]

B. Incidents Involving Private Security Guards

In December 2008, in New York City, Amtrak police arrested photographer Duane Kerzie for trespassing while he was taking pictures from the train platform.[64] Kerzie was taking pictures in an attempt to win Amtrak's annual photo contest entitled "Picture Our Trains."[65] The winner of the contest receives a grand prize of one thousand dollars in travel vouchers and the winning photo is published in Amtrak's annual calendar.[66] Amtrak security guards approached Kerzie with a black Labrador Retriever and instructed Kerzie to allow the dog to sniff his bag.[67] Kerzie complied and the dog sniffed his bag for explosives.[68] The security guards next asked for his ID and to see the photos. After viewing the photos, the guards demanded that Kerzie delete them.[69] Kerzie said "absolutely not," and the security guard replied that "it [is] illegal to photograph trains."[70] When Kerzie asked: "[W]here is the sign that says that?,"[71] the security guards

57. *Id.*
58. *Id.*
59. *Id.*
60. *Id.*
61. *Id.*
62. *Id.*
63. *Id.* Miller was successful in his appeal of the charge of resisting arrest without violence. *Id.*
64. Carlos Miller, *Amtrak Photo Contestant Arrested by Amtrak Police in NYC's Penn Station*, PHOTOGRAPHY IS NOT A CRIME, (Dec. 27, 2008, 2:29 AM), http://carlosmiller.com/2008/12/27/amtrak-police-arrest-photographer-participating-in-amtrak-photo-contest/.
65. *Id.*
66. *Id.*
67. *Id.*
68. *Id.*
69. *Id.*
70. *Id.*
71. *Id.*

pulled out handcuffs and arrested him on the spot.[72] Police placed Kerzie in a cell and later charged him with trespassing.[73]

In July 2008, Marilyn Parver, a 56-year-old grandmother, was taken off a JetBlue flight in handcuffs when she refused to delete a video she filmed of a "minor altercation between [two] passengers."[74] When she refused, the airline "threatened to blacklist her and accused her of interfering with a flight crew, which is a federal crime."[75] Two police officers, a TSA agent, and a JetBlue Airways representative escorted Parver off the flight.[76] She was taken to the Las Vegas Metropolitan Police Department, where police eventually released her with no charges.[77]

In October 2009, in Chicago, Illinois, an amateur photographer filmed a House of Blues security guard repeatedly pushing a female concertgoer to the ground.[78] The security guard took away the female's camera after she shot a picture of him.[79] The video captures the female attempting to take her camera back and the security guard violently shoving her and repeatedly pushing her to the ground.[80] The police arrested the security guard and charged him with misdemeanor battery.[81]

In October 2007, security guards removed Reza Michael Farhoodi, a University of Maryland student, from his seat in FedEx Stadium during a Washington Redskins game and questioned him about taking shots of the game and his friends with his SLR camera.[82] Security officers confiscated his camera and told him that he could pick it up at Guest Services after the game.[83] He was told that "professional" cameras were not permitted without press credentials, even though the printed rules did not distinguish between "professional" and "amateur" cameras.[84] Only hours later, a team executive

---

72. *Id.*

73. *Id.*

74. Christopher Elliot, *Grandmother Arrested After Refusing to Delete JetBlue Fight Video*, ELLIOT BLOG (Aug. 14, 2008), http://www.elliott.org/blog/grandmother-arrested-after-refusing-to-delete-jetblue-fight-video/.

75. *Id.*

76. Aaron Royster, *Woman Detained by Airline Over Video*, KINGMAN DAILY MINER (Aug. 7, 2008, 6:00 AM), http://www.kingmandailyminer.com/main.asp?SectionID=1&subsectionID=1&articleID=16860.

77. *Id.*

78. Steve Bryant, *HoB Security Guard Assaulted Young Girl: Cops*, NBC CHICAGO (Oct. 4, 2009, 12:14 PM), http://www.nbcchicago.com/news/local-beat/house-of-blues-hanson-assault-camera-64210757.html.

79. *Id.*

80. *Id.*

81. *Id.*

82. Marc Fisher, *At FedEx Field, an Eventual Victory for Shutterbugs*, THE WASHINGTON POST, Sept. 25, 2007, http://www.washingtonpost.com/wp-dyn/content/article/2007/09/24/AR2007092401521.html?sub=new.

83. *Id.*

84. *Id.*

called and apologized.[85] The executive explained that the ushers and officers were wrong for treating him as they did.[86]

C. Common Themes

The first major theme that emerges from all or most of these incidents is the use of broad statutes with sweeping language to arrest, charge, and prosecute photographers.[87] These statutes include, but are not limited to: loitering, disorderly conduct, assault, obstruction of justice, failure to obey police orders, disturbing the peace, provoking a riot, and resisting a police officer.[88] Eve Burton, Vice President and General Counsel of the Hearst Corporation, published an article in *Communications Lawyer* examining law enforcement officials' use of criminal statutes and tactics to limit newsgathering.[89] In the article, Burton argues that "[t]he criminal cases that present the greatest threat to a strong press [may be] insidious efforts by local police departments to curtail lawful newsgathering activities through the use of state criminal 'disorderly conduct,' 'assault,' and 'obstruction of justice' statutes."[90] She notes that these cases generally "stem from press coverage of crimes, accidents, and public appearances by political figures" and "are brought in nearly every state."[91] For example, it took nearly two years for California government officials to drop charges against a photographer for "interfering with a police officer" when he took pictures of police assaulting gang members.[92]

Another trend in many of these cases is police interference with photographers based on the all-encompassing notion of "national security."[93] Inquiries into which "national security" law they have violated often yield poor outcomes. Keith Garsee, a Los Angeles resident, described what happened when he took a picture with his camera phone while waiting to board a subway.[94] Garsee explained that after he took the picture, a subway employee stated: "Hey! It's against the 9-11 Law to take pictures

---

85. *Id.*

86. *Id.*

87. *See* Eve B. Burton, *Where Are All the Angry Journalists? The Use of Criminal Statutes and Tactics to Limits Newsgathering*, COMM. LAW, Summer 1998, at 19, 19.

88. *See id.* at 19–21. *See generally* PHOTOGRAPHY IS NOT A CRIME, http://photography isntacrime.com/ (last visited Jan. 25, 2011) (providing multiple examples of photographers charged under a variety of statutes for taking photographs).

89. Burton, *supra* note 87.

90. *Id.* at 20–21.

91. *Id.* at 21.

92. *Id.*

93. *See, e.g.*, Keith Garsee, *Orweillian Los Angeles*, KEITH GARSEE'S MYSPACE BLOG (May 14, 2008, 3:13 PM), http://blogs.myspace.com/index.cfm?fuseaction=blog.view &friendID=71473815&blogID=394235689.

94. *Id.*

down here man!"[95] Garsee asked the subway employee to which law he was referring. The subway employee asked Garsee if he was a lawyer, to which Garsee replied "no." The subway employee then stated, "No pictures. You could be a terrorist. Very strict!"[96] Garsee said that shortly following the altercation with the subway employee, a woman announced over the intercom: "Attention to the gentleman in the plaid shirt: you are not allowed to take photographs in the Subway. You will be arrested if you continue to take photos and harass the metro worker."[97] After getting off the subway, Garsee contacted the sheriff's station and spoke with a deputy who told him that "there is no such law."[98]

Ignored in many of these cases is the fact that a Google image search can reveal a multitude of images of many of these places, some of which are likely similar to those the photographers have captured (or have attempted to capture) with their lenses.[99] In fact, Google Earth allows one to retrieve photos of nearly any address in the country from a variety of angles and distances.[100] Given this broad, public access to images of public buildings, transportation stations, and the like, it is not entirely clear why law enforcement officials, or those who write their policy manuals, are convinced that limiting photography enhances national security.[101]

III. PHOTOGRAPHERS' RIGHTS—THE EXISTING LEGAL FRAMEWORK

One of the more troubling things about the incidents discussed in Section II is that no law exists prohibiting the photographers' behavior.[102] As a general rule, subject to only a few exceptions, photographers have the right to take pictures in public places and in private places opened to the public.[103] This section discusses the existing legal regime surrounding the right to take photographs.

---

95. *Id.*
96. *Id.*
97. *Id.*
98. *Id.*
99. *See* GOOGLE IMAGES, http://www.google.com/imghp (last visited Jan. 25, 2011).
100. *See* GOOGLE EARTH, http://www.google.com/earth/index.html (last visited Jan. 25, 2011).
101. *See* Bruce Schneier, *The War on Photography*, SCHNEIER ON SECURITY (June 5, 2008, 6:44 AM), http://www.schneier.com/blog/archives/2008/06/the_war_on_phot.html.
102. *See* Note, *Privacy, Photography, and the Press*, 111 HARV. L. REV. 1086, 1088 (1998).
103. *See id.*

A. Public Places

1. The Right to Photograph People in Public Places

Photographers generally have a right to take pictures of others in public places.[104] This is the case even if the subject of the photograph has not authorized the picture.[105] "[C]ourts consistently have upheld the rights of photographers to take unauthorized photographs of others in or from public places."[106] Dean Prosser explains:

> On the public street, or in any other public place, the plaintiff has no right to be alone, and it is no invasion of his privacy to do no more than follow him about. Neither is it such an invasion to take his photograph in such a place, since this amounts to nothing more than making a record, not differing essentially from a full written description, of a public sight which any one present would be free to see.[107]

Tort law therefore assumes that persons implicitly consent to being photographed upon leaving the confines of their homes.[108] However, both tort and criminal law recognize a few exceptions to this general rule and place limits on the methods photographers may use to obtain photographs of people.[109]

Four distinct torts exist that protect the privacy interests of the individual: (1) intrusion upon the seclusion or solitude of another, (2) public disclosure of private facts, (3) publicity that places another in a false light, and (4) appropriation of another's name or likeness for one's own advantage.[110] Generally, a privacy tort action against a photographer taking pictures in public places only exists where a photographer has failed to act within the bounds of common decency and respect for others.[111] For example, a claim for intrusion upon seclusion would lie where a sudden gust of wind blows a woman's skirt over her head and a photographer

---

104. *See id.* at 1089 (noting the law's general "assumption that a person who leaves the confines of a private location implicitly consents to being photographed").

105. *Id.*

106. *Id.* at 1088 (citing Fogel v. Forbes, Inc., 500 F. Supp. 1081, 1087 (E.D. Pa. 1980); Pemberton v. Bethlehem Steel Corp., 502 A.2d 1101, 1116–17 (Md. Ct. Spec. App. 1986); Forster v. Manchester, 189 A.2d 147, 150 (Pa. 1963)); *see also* Durant v. State, 188 P.3d 192, 195 (Okla. Crim. App. 2008).

107. William L. Prosser, *Privacy*, 48 CAL. L. REV. 383, 391–92 (1960) (footnotes omitted).

108. Note, *supra* note 102, at 1089.

109. *Id.* (noting the existence of various privacy torts as well as criminal statutes such as harassment).

110. RESTATEMENT (SECOND) OF TORTS § 652A (1976).

111. *See* RESTATEMENT (SECOND) OF TORTS §§ 652A–E (1976).

immediately begins snapping pictures.[112] Likewise, a photographer may not use a high-powered lens to stand on a public sidewalk and photograph the inside of a person's bedroom through a crack in the blinds.[113]

In addition to these limits, several states have civil and criminal statutes prohibiting harassment.[114] For example, California's harassment statute refers to "a knowing and willful course of conduct directed at a specific person which seriously alarms, annoys, or harasses the person, and which serves no legitimate purpose."[115] New York enacted a statute stating "that a person is guilty of harassment . . . when he or she possesses the 'intent to harass, annoy, or alarm another person,' and 'follows a person in or about a public place' or 'engages in a course of conduct . . . which alarms or seriously annoys' another person."[116] Thus, in obtaining a photograph of a person in a public place, photographers in some states are prohibited from using intrusive tactics like those described above.

Criminal penalties may also result from interfering with police investigations of crimes or accidents.[117] In most states, photographers must remain a certain minimum distance from accidents and police investigations.[118] Courts have routinely upheld these limitations on the ground that police need room to do their job, and a mob of photographers and flashes will almost certainly impede their progress.[119]

Courts have also upheld prohibitions on taking pictures in areas that impede traffic.[120] In *Siegman v. District of Columbia*, the District of Columbia Court of Appeals upheld a police regulation mandating that "No person licensed [thereunder] should impede traffic while engaged in taking photographs, nor remain longer than five minutes at any one location on the streets, sidewalks, or other public places."[121] While the state (or District, in the case of *Siegman*) might have a valid interest in restricting photographers' rights in the name of security or maintaining order, these types of regulations must be narrowly tailored to prevent imposing restrictions on photographers greater than those necessary to prevent

112. RESTATEMENT (SECOND) OF TORTS § 652B cmt. c, illus. 7 (1976).

113. *See id.* § 652B cmt. b, illus. 2.

114. Note, *supra* note 102, at 1089 (noting existence of harassment laws).

115. *Id.* at n.26 (citation omitted).

116. *Id.* (citation omitted).

117. *See, e.g.*, N.Y. PENAL Law § 195.05 (McKinney 2010).

118. *See, e.g., id.*

119. *See, e.g.*, Decker v. Campus, 981 F. Supp. 851, 857 (S.D.N.Y. 1997) (holding that "where the fact of physical interference with an official function is clear as a matter of law, an officer's decision to arrest an individual for obstructing governmental administration is generally reasonable"); State v. Lashinsky, 404 A.2d 1121, 1130 (N.J. 1979) (finding that the action of an officer in ordering a photographer to move away from an accident scene was "plainly reasonable in objective terms because [the photographer's conduct caused] an actual interference" in the officer's work).

120. *See, e.g.*, Siegman v. District of Columbia, 48 A.2d 764, 767 (D.C. 1946).

121. *Id.* at 764, 767.

impeding traffic.[122] For example, in *Connell v. Town of Hudson*, the federal district court in New Hampshire noted that police officers violated a news photographer's First Amendment rights at an automobile accident scene to the extent that the restrictions imposed upon him were greater than those necessary to prevent his unreasonable interference with the police and emergency functions.[123]

In sum, a photographer's right to take pictures of people in public places is fairly broad under existing federal and state laws.[124] "People" includes all people—children, security guards, police officers, people committing crimes, people being arrested, celebrities, etc.[125] Tort law assumes that when a person leaves his or her home, he or she implicitly consents to being photographed.[126] The four torts protecting individuals' privacy interests usually require some type of clearly intrusive or obviously inappropriate behavior on behalf of the photographer.[127] This is also the case with civil and criminal harassment.[128] Lastly, photographers are prohibited from getting too close to accident scenes or investigations when photographing police, victims, or arrestees, or they run the risk of being charged pursuant to criminal statutes such as obstruction of justice.[129]

2. The Right to Photograph Structures Visible from Public Places

Courts have consistently upheld the right to photograph buildings visible from public places.[130] In 1990, Congress passed the Architectural Works Copyright Protection Act, establishing a new category of copyright protection for works of architecture.[131] However, Congress specifically limited the protection afforded architectural works with an amendment protecting the right to photograph buildings visible from public places.[132]

---

122. *See, e.g.*, Connell v. Town of Hudson, 733 F. Supp. 465, 469 (D.N.H. 1990).

123. *Id.*

124. *See* Prosser, *supra* note 107, at 391–92.

125. *See* BERT KRAGES, LEGAL HANDBOOK FOR PHOTOGRAPHERS: THE RIGHTS AND LIABILITIES OF MAKING IMAGES 26 (Michelle Perkins ed., 2nd ed.) (2007).

126. Note, *supra* note 102, at 1089.

127. *See* RESTATEMENT (SECOND) OF TORTS §§ 652A–E.

128. *See* Note, *supra* note 102, at 1089 n.26.

129. *See, e.g.*, State v. Lashinsky, 404 A.2d 1121, 1128 (N.J. 1979).

130. *See, e.g.*, R.M.S. Titanic, Inc. v. Haver, 171 F.3d 943, 970 (4th Cir. 1999) (footnote omitted) ("[A] property right does not normally include the right to exclude viewing and photographing of the property when it is located in a public place."); Landrau v. Solis Betancourt, 554 F. Supp. 2d 102, 113 (D.P.R. 2007) (noting that the Architectural Works Copyright Protection Act, 17 U.S.C. § 120(a) (2006), allows the taking and publishing, without the architect's consent, of photographs of buildings located in or visible from public places).

131. 17 U.S.C. § 102(a)(8) (2006).

132. *See* 17 U.S.C. § 120(a) (2006). The amendment states,

In copyright suits involving the right to photograph buildings, courts have taken this amendment seriously, typically ruling in favor of photographers.[133]

"Buildings" include both residential and commercial buildings.[134] Bridges, industrial facilities, public utilities, transportation facilities (e.g., airports), and almost all other structures visible from public places fall within this definition.[135] However, a few exceptions do exist. Commanders of military installations can prohibit photographs of specific areas when they deem it necessary to protect national security.[136] Additionally, the U.S. Department of Energy can prohibit photography of designated nuclear facilities although the publicly visible areas of nuclear facilities are usually not designated as such.[137] Outside of these narrow exceptions, the right to photograph buildings and other structures is virtually unlimited under existing law.[138]

B. Privately Owned Places Open to the Public

The law governing the right to take photographs in privately owned places open to the public is relatively straightforward. As discussed earlier, there are virtually no limits on the right to photograph privately owned structures that are visible from public places.[139] Property owners have no right to prohibit others from photographing their property from other locations.[140] However, property owners *do* have a right to prohibit photography occurring *on* their property, even if it is open to the public.[141] If no signs are posted prohibiting photography in areas such as shopping malls, private museums, amusement parks, restaurants, train stations, and

---

> The copyright in an architectural work that has been constructed does not include the right to prevent the making, distributing, or public display of pictures, paintings, photographs, or other pictorial representations of the work, if the building in which the work is embodied is located in or ordinarily visible from a public place.

Id.

133. *See, e.g.*, Leicester v. Warner Bros., 232 F.3d 1212, 1228–29 (9th Cir. 2000) ("In other words, if you want to copyright a building as constructed and thereby prevent others from constructing buildings that copy your design, you have to permit people to take, display and distribute pictures of your building without limitation.").

134. *See* KRAGES, *supra* note 125, at 18–19.

135. *Id.* at 14.

136. *Id.* at 43–44.

137. *Id.* at 44.

138. *See, e.g.*, *R.M.S. Titanic Inc.*, 171 F.3d at 970.

139. *Id.*

140. *Id.*

141. *See* KRAGES, *supra* note 125, at 18–19.

airports, it is probably safe to assume that it is allowed.[142] Further, if no signs are posted, photographers are free to take pictures of people or things on the property,[143] subject only to the limitations described above (e.g. privacy torts, harassment, obstruction of justice for interfering with an investigation, accident, arrest, etc.).[144] Even if no signs are posted, property owners or operators may still request that one stop taking pictures.[145] If a photographer ignores these requests, he or she may be charged with trespassing.[146] Property owners are not, however, permitted to confiscate one's camera or film.[147]

C. Observations

A few things become apparent after examining the existing legal framework surrounding photography rights. First, there are very few actual limits on the right to take pictures in public.[148] Second, existing limits seem to be aimed at the most obnoxious and intrusive behavior, such as taking pictures underneath a woman's skirt or repeatedly harassing someone after being asked to stop taking pictures.[149] Third, no post-9/11 laws specifically prohibit the right to take pictures in public.[150] With so few "on the books" limitations, one may wonder how and why photographers such as those discussed in Part II are continually questioned, arrested, and charged for taking pictures in public places.

In many instances, law enforcement officials assert authority pursuant to the following: (1) broadly worded criminal statutes that were never intended to apply to photographers under the existing circumstances;[151] (2)

---

142. *Id.* at 19.

143. *Id.*

144. *See supra* text accompanying notes 109–29.

145. *See* KRAGES, *supra* note 125, at 18–20.

146. *Id.* at 18–19 ("[T]here is no general legal right of access to private property for the purpose of taking photographs, which means that photographers must obey the same laws that apply to the general public. Because private property owners have the right to exclude others from their property and to limit the activities of those they allow to enter, photographers face liability for trespass if they enter . . . .").

147. *Id.* at 22.

148. *See, e.g.*, *Leicester*, 232 F.3d at 1228–29; *R.M.S. Titanic, Inc.*, 171 F.3d at 970; *Landrau*, 554 F. Supp. 2d at 113.

149. *See* KRAGES, *supra* note 125, at 25–30 ("Despite the importance that society places on personal privacy, the law imposes relatively few restrictions on photographing people.").

150. *See* Bruce Schneier, *Are Photographers Really a Threat?*, THE GUARDIAN (June 5, 2008), http://www.guardian.co.uk/technology/2008/jun/05/news.terrorism ("There's nothing in any post-9/11 law the restricts your right to photograph.").

151. Michael K. Powell, *The Public Interest Standard: A New Regulator's Search for Enlightenment*, 16 COMM. LAWYER: J. MEDIA, INFO. & COMM. LAW 19, 19 (1998); *see also* Burton, *supra* note 87, at 19–21.

"national security" or "9/11" laws;[152] or (3) nothing at all (i.e., "it's just the law" or "taking pictures here is illegal"). Often, charges are eventually dropped and apologies are issued,[153] but at that point, the damage has already occurred. It is impossible to recreate the newsworthy events a journalist could have captured had he or she not been handcuffed. Likewise, it is impossible to "undelete" a memory card with days', months', or even years' worth of pictures. Neither dropping charges nor issuing apologies compensates a photographer for the embarrassment and possible harm to his reputation experienced when a crowd of people sees him surrounded by blue lights and shoved into a police car. Moreover, a photographer whose rights have been abused has very few legal remedies,[154] and most of the remedies that are available are either inadequate or are totally impractical when the costs of litigation are considered.[155] The next section explores these remedies and explains why each is inadequate.

IV. REMEDIES FOR VIOLATIONS COMMITTED BY LAW ENFORCEMENT OFFICIALS IN PUBLIC PLACES

A. Declaratory Relief

The Uniform Declaratory Judgments Act, which has been adopted with slight modifications in the vast majority of states, provides that "[c]ourts . . . shall have power to declare rights, status, and other legal relations whether or not further relief is or could be claimed."[156] The Federal Declaratory Judgment Act provides the following:

> In a case of actual controversy within its jurisdiction . . . any court of the United States, upon the filing of an appropriate pleading, may declare the rights and other legal relations of any interested party seeking such declaration, whether or not further relief is or could be sought. Any such declaration shall have the force and effect of a final judgment or decree and shall be reviewable as such.[157]

In certain situations, declaratory judgment statutes can provide relief to a plaintiff whose rights have been violated within the context of photography. For example, a photographer who is charged pursuant to a broadly worded "obstruction of justice" statute may bring an action seeking a declaration that the statute is unconstitutionally vague or broad or that the behavior he or she was engaged in is lawful. Additionally, if a law

---

152. *See, e.g.*, Garsee, *supra* note 93.
153. *See supra* notes 82–86 & 90–92 and accompanying text.
154. *See* discussion *infra* Part IV.
155. *Id.*
156. UNIF. DECLARATORY JUDGMENTS ACT, G.L. § 1 (1922).
157. 28 U.S.C. § 2201(a) (2006).

enforcement officer threatens a photographer with charges for lawful behavior, and that threat is sufficiently real, the photographer may seek from a court a declaration of his or her rights or confirmation that the behavior at issue is lawful.

Certain obstacles and limitations are associated with declaratory relief. First, a court will not award declaratory relief unless the claim is sufficiently "immediate and real."[158] Raymond Beauchamp provides the following guidance:

> The Supreme Court has stated that declaratory relief is appropriate when "a refusal on the part of the federal courts to intervene . . . may place the hapless plaintiff between the Scylla of intentionally flouting state law and the Charybdis of forgoing what he believes to be constitutionally protected activity. To determine if a claim is sufficiently immediate and real, "[t]he [Supreme] Court spread the justiciability question along a continuum ranging between 'a general threat by officials to enforce those laws which they are charged to administer' and 'a direct threat of punishment against a named [party] . . . for a completed act'" with those closer to the direct threat more likely to be an actual controversy.[159]

A photographer merely *threatened* with charges may experience difficulty demonstrating a "direct threat of punishment." If he or she is not able to convince a court that the threat is sufficiently immediate, he or she is forced to decide between foregoing what he or she believes to be constitutionally protected behavior and potentially being charged with a crime.

Even if a plaintiff establishes a justiciable claim, it may be difficult to convince a judge to declare the statute at issue (for example, "obstruction of justice") unconstitutionally vague or broad. Law enforcement officials could argue that the statute is only applied when a person comes within a certain number of feet from the arrest and interferes with police duties. Even if the journalist was much further from the arrest—which sophisticated cameras make possible—it is the law enforcement official's word against the journalist's. A judge may be hesitant to strike down a statute that officers can "legitimately" use to prevent the press from interfering with the successful performance of their duties.

Even if a judge were to declare the statute unconstitutionally vague or broad and drop the charges, the journalist has still has suffered a substantial amount of harm. She has been publicly humiliated by being handcuffed and placed in a police car. Additionally, she may have lost the opportunity to cover a breaking news story. Even if a court declares the statute unconstitutional, the journalist will still be responsible for the time commitment and costs associated with litigation. Further, if the statute is

158. Raymond W. Beauchamp, *England's Chilling Forecast: The Case for Granting Declaratory Relief to Prevent English Defamation Actions from Chilling American Speech*, 74 FORDHAM L. REV. 3073, 3103 (2006).

159. *Id.*

eventually rewritten to define "obstruction of justice" more narrowly, say to include coming within ten feet of an arrest or investigation, not much has changed. The scenario could repeat itself with the same outcome—the police officer's word against the journalist's.

Lastly, declaratory judgment represents a limited remedy in that it provides a plaintiff only with a declaration of his or her rights or the constitutionality of a statute, not pecuniary damages. A declaration of one's rights will allow a plaintiff to resume his or her constitutionally protected activities and result in charges being dropped; however, it will not compensate a photographer or journalist for the loss of news stories, damage to person or property, or deleted pictures.

B. Injunctive Relief

An injunction is an equitable remedy in the form of a court order, whereby a party is required to do, or to refrain from doing, certain acts.[160] To secure an injunction, a plaintiff must meet the Article III "case or controversy" requirement.[161] The Ninth Circuit set the following standard:

> The "irreducible minimum" demanded of a proper plaintiff by Article III's constitutional demands . . . requires that a plaintiff show he has "personally . . . suffered some actual or threatened injury as a result of the putatively illegal conduct of the defendant," that can be "fairly" traced to the defendant's challenged conduct, and which "is likely to be redressed by a favorable decision."[162]

Under limited circumstances, an injunction may provide a remedy for a plaintiff within the context of photography and newsgathering. Suppose the plaintiff is a well-known journalist in a city who has covered instances of police misconduct in the past and has had a number of altercations with police while covering stories. He has been arrested for obstruction of justice while photographing a political rally (charges were dropped) and threatened with charges a number of other times. The plaintiff believes that his First and Fourth Amendment rights were violated and will be violated in the future.

To succeed in a claim for injunctive relief, the plaintiff would need to show that his First or Fourth Amendment rights were violated.[163] This requires showing that officers lacked probable cause to arrest him and thus

---

160. *See* Thompson v. Lantz, No. 3:04cv2084, 2008 WL 762465, at *1 (D. Conn. Mar. 20, 2008)

161. *See* LaDuke v. Nelson, 762 F.2d 1318, 1323 (9th Cir. 1985).

162. *Id.* The court further noted that "at least when injunctive relief is sought, litigants must adduce a 'credible threat' of recurrent injury." *Id.*

163. *See* Worthington v. Kenkel, 684 N.W.2d 228, 232 (Iowa 2004).

violated his Fourth Amendment rights.[164] He would also need to show that his behavior did not violate any legitimate laws (i.e., he observed distance limitations, etc.) and is protected by the First Amendment.[165] The plaintiff must demonstrate that his arrest and the officers' many threats to arrest him represent a "pattern of police misconduct."[166] Lastly, the plaintiff must show that he is likely to be threatened, arrested, or both in the future for newsgathering activities that are protected by the Constitution.[167]

Securing an injunction under these circumstances will be difficult. First, police officers will likely take the position that the plaintiff was in fact interfering with the execution of their duties by being too close or interfering with the police officers' control over the situation.[168] It will be the officers' word against the journalist's, and a judge may show deference to the officers. Second, showing a pattern of illicit law enforcement behavior may be difficult. Cases in which the court has found such patterns have included extensive factual findings over extended periods of time.[169] In any case, showing a pattern of illicit behavior represents a fact-specific inquiry[170] that can translate into costly and time-consuming litigation. Additionally, injunctive relief, like declaratory relief, is an equitable remedy and may not provide compensatory damages.

C. 42 U.S.C. § 1983 Action

42 U.S.C. § 1983 provides that

> Every person who, under color of any statute, ordinance, regulation, custom, or usage, of any State or Territory or the District of Columbia, subjects, or causes to be subjected, any citizen of the United States or other person within the jurisdiction thereof to the deprivation of any

---

164. *See* Stewart v. Abraham, 275 F.3d 220, 228 (3d Cir. 2001) (discussing injunctions within the context of the Fourth Amendment).

165. *See Worthington*, 684 N.W.2d at 232.

166. *See* Thomas v. Cnty. of Los Angeles, 978 F.2d 504, 508 (9th Cir. 1992) ("[T]he Supreme Court requires a showing of an intentional and pervasive pattern of misconduct in order to enjoin a state agency.").

167. *See* Haney v. Miami-Dade Cnty., Nos. 04-20516-CIV-JORDAN, 04-20516-CIV-BROWN, 2004 WL 2203481, at *2–3 (S.D. Fla. Aug. 24, 2004). The court held that an injunction was proper and recognized that "when the threatened acts that will cause injury are authorized or part of a policy, it is significantly more likely that the injury will occur again." *Id.* (quoting 31 Foster Children v. Bush, 329 F.3d 1255, 1266 (11th Cir. 2003)).

168. *See* State v. Taylor, 118 A.2d 36, 49 (N.J. Super. Ct. App. Div. 1955) (holding defendant guilty for interference with an officer in lawful discharge of his duties while using loud and offensive language).

169. *See, e.g.*, LaDuke v. Nelson, 560 F. Supp. 158, 160 (E.D. Wash. 1982); Allee v. Medrano, 416 U.S. 802, 805–09 (1974).

170. *Id.*

> rights, privileges, or immunities secured by the Constitution and laws, shall be liable to the party injured in an action at law . . . .[171]

Section 1983 provides a federal remedy for individuals whose federal constitutional rights are violated by persons acting under the color of state law.[172] In most cases, private security guards will not be considered persons acting under the color of state law for the purposes of this statute.[173] However, exceptions exist and will be covered in Part V.[174]

There are three major hurdles that a § 1983 plaintiff within the photography context must clear: (1) establishing a constitutional violation,[175] (2) the doctrine of qualified immunity,[176] and (3) proving damages.[177] It is noted that since § 1983 applies only to persons acting under color of state (not federal) law, references to the First and Fourth Amendment assume incorporation of these amendments into the Fourteenth Amendment through its Due Process clause.[178] The following section examines these three hurdles using a variety of hypothetical scenarios.

1. Establishing a Constitutional Violation

Suppose Person A is standing on a sidewalk photographing a major bridge. A security guard approaches her and informs her that she is not allowed to take pictures of the bridge for reasons of "national security." Person A is convinced she is doing nothing wrong and continues to take pictures. The security guard threatens to call the police and eventually does to report the "suspicious activity." The police come, blue lights flashing, and interrogate Person A for thirty minutes. They finally leave after forcing her to delete her pictures by threatening to place her on an FBI watch list. Person A refrains from taking any more pictures of bridges or buildings for fear of similar consequences.[179] Person A brings a § 1983 claim for violation of her First Amendment rights, seeking damages for the embarrassment and humiliation she experienced while the police were questioning her, as well as the loss of all the pictures on her camera.

As stated above, § 1983 only applies to persons acting under color of state law, so Person A probably will not be able to name the private security

---

171. 42 U.S.C. § 1983 (2006).

172. *Id.*

173. *But see* discussion *infra* Part V.A (discussing exceptions where security guards were found to be acting under color of state law).

174. *Id.*

175. *See* Montefusco v. Nassau Cnty., 39 F. Supp. 2d 231, 242 n.7 (E.D.N.Y. 1999).

176. Smith v. Kenny, 678 F. Supp. 2d 1093, 1105 (D.N.M. 2009) (citations omitted).

177. United States v. Henry, 447 U.S. 264, 269 (1980).

178. U.S. CONST. amend. XIV, § 1.

179. *See* Portland incident *supra* Part II.A. (providing the basis for this hypothetical).

guard as a defendant.[180] Thus, Person A is only able to sue the police officers. Person A could assert that her First Amendment right to freedom of expression was violated.[181] However, since "images . . . must communicate some idea in order to be protected under the First Amendment," Person A's claim is unlikely to hold up.[182] Even when plaintiffs have claimed that they plan to use photographs they have taken for expressive and transformative purposes, courts have refused First Amendment protection in these instances.[183] Therefore, Person A must allege something other than a First Amendment violation to recover under § 1983.

Person B, a journalist for a small newspaper, is riding in his friend's car when his friend is pulled over for no apparent reason. The officer never articulates a reason for stopping the vehicle. He asks Person B's friend whether he has been drinking and whether there are weapons or drugs in the vehicle. Despite the friend's insistence that he has not been drinking and that there are no drugs or weapons in the vehicle, the officer continues to question him aggressively about where they have been and where they are going. Outraged at the inappropriate and invasive nature of the officer's questioning, Person B begins recording the exchange on his phone. After making both get out of the car and conducting a search of the vehicle, the officer notices that Person B has been recording the entire incident. He demands that Person B stop recording and delete the footage. Without doing either, Person B puts the phone into his pocket. The officer confiscates Person B's phone, deletes the video, and arrests him for failing to obey police orders. The officer lets Person B's friend drive away. Charges are eventually dropped, but Person B brings a § 1983 suit for violation of his First Amendment right to freedom of the press, among other rights.

Although Person B has a better chance at demonstrating a violation of freedom of the press than Person A (since he may be considered "press" capturing a potential news story),[184] his claim is still questionable. As Eric Ugland has noted, "courts have not reached a consensus about the shape that a definition [of the press] should take. The Supreme Court has provided little guidance, and the lower court approaches and statutory definitions are inconsistent and sometimes arbitrary."[185] It is not entirely clear whether a

---

180. *But see* discussion *infra* Part V.A. (discussing exceptions where security guards were found to be acting under color of state law).

181. *See* Montefusco v. Nassau Cnty., 39 F. Supp. 2d 231, 241 (E.D.N.Y. 1999) (citations omitted).

182. *See id.*

183. *Id.* at 242 n.7.

184. *See* Erik Ugland, *Demarcating the Right to Gather News: A Sequential Interpretation of the First Amendment*, 3 DUKE J. CONST. L. & PUB. POL'Y 113, 174–75 (2008).

185. *Id.* at 137. "As Justice White wrote in *Branzburg*, trying to define who is a

court would consider Person B a member of the "press"[186] for the purpose of First Amendment protection. Even if Person B is classified as press, courts are torn over the extent to which First Amendment protections extend to the newsgathering process.[187] The Supreme Court has recognized that "without some protection for seeking out the news, freedom of the press could be eviscerated."[188] While most courts agree that newsgathering does "qualify for First Amendment protection,"[189] this protection is qualified.[190] According to the Supreme Court of Ohio, for instance, "[t]he [F]irst [A]mendment of course does not immunize wrongful behavior simply because it is undertaken in the name of newsgathering."[191] One can see how police officers can claim "wrongful behavior" for any number of actions by the press an officer considers to be interfering with an investigation.

In Person B's case, any First Amendment newsgathering right he possesses may be trumped if a state statute exists prohibiting the recording of police investigations. In Massachusetts for example, a statute prohibits "interception of any . . . oral communication . . . ."[192] Under Massachusetts law, "the term 'interception' means to secretly hear, secretly record, or aid another to secretly hear or secretly record the contents of any wire or oral communication through the use of any intercepting device. . . ."[193] In *Commonwealth v. Hyde*, the defendant was prosecuted for secretly recording a police investigation during a routine traffic stop.[194] The dissenting justices of the Massachusetts Supreme Judicial Court argued that

---

journalist would 'present practical and conceptual difficulties of a high order.'" *Id.* at 136.

186. *Id.* at 137.

187. Professor Ugland writes:

> Over the past three decades, journalists have sought to broaden the definition of press freedom to protect newsgathering, arguing that if they are to serve the highest purposes of their profession, freedom of the press must encompass more than the right to publish what they know . . . . They have also challenged restrictions that intrude too deeply on . . . journalistic expression and investigative zeal . . . .
>
> Some courts have been sympathetic to these challenges, but many have rejected them, showing little patience for what judges often construe as media demands for "special rights."

Id. at 120–21.

188. Richmond Newspapers, Inc. v. Virginia, 448 U.S. 555, 576 (1980) (citing Branzburg v. Hayes, 408 U.S. 665, 681 (1972)).

189. *Branzburg*, 408 U.S. at 681.

190. Boddie v. Am. Broad. Co., 881 F.2d 267, 271 (Ohio 1989).

191. *Id.*

192. MASS. GEN. LAWS ch. 272, § 99 (2010).

193. *Id.* § 99(B)(4).

194. Commonwealth. v. Hyde, 750 N.E.2d 963, 969 (Mass. 2001).

"[t]he defendant's secret recording of the words of the police officers should be lawful, because such recording may tend to hold police officers accountable for improper behavior."[195]

In sum, Person B will have a difficult time convincing a court that his First Amendment right to freedom of the press was violated. First, it is uncertain whether Person B qualifies as press.[196] Second, it is not clear that a court would extend First Amendment protection to newsgathering under these circumstances.[197] Third, even if it did, the presence of a state statute prohibiting Person B's behavior will trump any First Amendment protections he may possess.[198]

In the context of photography, a § 1983 plaintiff will have more success establishing a violation of his or her Fourth Amendment rights. The Fourth Amendment provides that "[t]he right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated . . . but upon probable cause . . . ."[199] "[A] search ordinarily must be based on individualized suspicion of wrongdoing."[200] In order "[t]o prevail on a § 1983 claim under the Fourth Amendment based on an allegedly unlawful *Terry* stop,[201] a plaintiff first must prove that he was seized."[202] A police officer may approach an individual and ask questions without intruding on Fourth Amendment rights, because this would not be a seizure.[203] "[T]o determine whether a [police] encounter constitutes a seizure, a court must consider . . . whether a reasonable person . . . [would feel] free to decline the officers' requests or otherwise terminate the encounter."[204] If a court finds that a seizure occurred, a plaintiff must then demonstrate that the police officer lacked "reasonable suspicion of criminal activity."[205] A court will consider the

---

195. *Id.* at 969.

196. *See supra* text accompanying notes 184–86.

197. *See supra* text accompanying notes 187–91.

198. *See supra* text accompanying notes 192–95.

199. U.S. CONST. amend. IV.

200. Relford v. Lexington-Fayette Urban Cnty. Gov't, 390 F.3d 452, 458 (6th Cir. 2004) (citing Chandler v. Miller, 520 U.S. 305, 313 (1997)).

201. Terry v. Ohio, 392 U.S. 1, 30–31 (1968) ("[W]here a police officer observes unusual conduct which leads him reasonably to conclude in light of his experience that criminal activity may be afoot and that the persons with whom he is dealing may be armed and presently dangerous, where in the course of investigating this behavior he identifies himself as a policeman and makes reasonable inquiries, and where nothing in the initial stages of the encounter serves to dispel his reasonable fear for his own or others' safety, he is entitled for the protection of himself and others in the area to conduct a carefully limited search of the outer clothing of such persons in an attempt to discover weapons which might be used to assault him.").

202. Brown v. City of Oneonta, 221 F.3d 329, 340 (2d Cir. 2000).

203. United States. v. Lee, 916 F.2d 814, 819 (2d Cir. 1990).

204. Florida v. Bostick, 501 U.S. 429, 439 (1991).

205. United States v. Gray, 213 F.3d 998, 1000 (8th Cir. 2000).

"totality of the circumstances' of each case to see whether the detaining officer has a 'particularized and objective basis' for suspecting legal wrongdoing."[206] In order to actually arrest an individual without a warrant, police officers must have probable cause to believe that the individual committed a crime.[207]

In many of the incidents described in this article, a seizure by police officers did occur.[208] In some of the incidents, police arrested persons taking pictures.[209] In others, multiple officers arrived on the scene and questioned those taking pictures for extended periods of time.[210] A court could reasonably find in many, if not most, of the incidents, that the person being questioned did not feel "free to decline the officers' requests or otherwise terminate the encounter."[211] The bigger issue will be "whether the detaining officer ha[d] a 'particularized and objective basis' for suspecting legal wrongdoing."[212]

Simply taking pictures of tall buildings or major bridges—nothing more—certainly does not give rise to a reasonable suspicion of criminal activity.[213] No law exists prohibiting taking pictures of these structures,[214] so the "criminal activity," from the officer's perspective, is presumably the plotting of a terrorist attack.[215] But millions of people take pictures of tall buildings all of the time, so before seizing a person, an officer's reasonable suspicion must be based on some other facts that indicate the plotting of a terrorist attack.[216] In the Portland incident described above, the police had no other facts yet questioned the two girls for an extended period of time.[217] The police eventually contacted the FBI, describing the girls as "two Middle Eastern-looking teenagers taking pictures near Portland bridges."[218] A court hearing these facts would likely conclude that the officers lacked reasonable suspicion to seize them.[219]

---

206. United States. v. Arvizu, 534 U.S. 266, 273 (2002) (citation omitted).

207. *See* United States v. McClain, 444 F.3d 556, 562–63 (6th Cir. 2005).

208. *See* incidents *supra* Part II.

209. *See supra* text accompanying notes 17–24, 34–39, 55–63.

210. *See supra* text accompanying notes 11–16, 25–33, 40–49.

211. Florida. v. Bostick, 501 U.S. 429, 439 (1991).

212. *See* Harris v. State, 806 A.2d 119, 126 (Del. 2002) (quoting *Arvizu*, 534 U.S. at 273).

213. *See* United States v. McClain, 444 F.3d 556, 562 (6th Cir. 2005).

214. *See supra* notes 130–38 and accompanying text.

215. *See, e.g.*, incidents described *supra* notes 11–16.

216. Ortega Melendres v. Arpaio, 598 F. Supp. 2d 1025, 1033 (D. Ariz. 2009)

217. *See supra* notes 11–16 and accompanying text.

218. *See supra* note 15 and accompanying text.

219. *See Harris*, 806 A.2d at 127–28. "[A]n 'inchoate and unparticularized suspicion or hunch' of experienced police officers is insufficient to support a finding of reasonable suspicion as a matter of law." *Id.* (quoting Reid v. Georgia, 448 U.S. 438, 441 (1980)).

As stated earlier, if an officer makes an arrest without a warrant, he or she must have probable cause to believe the suspect committed a crime.[220] In most cases, arresting someone without a warrant or probable cause constitutes an unreasonable seizure in violation of the Fourth Amendment.[221] For a seizure to be reasonable, the behavior the suspect is believed to be engaging in must be criminal pursuant to some statute, law, or regulation.[222] Arresting someone in the name of "national security" for engaging in conduct that could hardly be construed as suspicious is not enough. A § 1983 plaintiff in such a case could successfully demonstrate a Fourth Amendment violation.

Suppose officers are making a forceful arrest on a street corner. A man walking by immediately begins snapping photographs of the incident. He is told to stop taking pictures. He continues to take pictures from a distance. Other people begin noticing the chaos and gather around the scene. The man is arrested for provoking a riot. Assume provoking a riot is prohibited by statute. Did the police have probable cause to believe that the man provoked a riot by taking pictures of an already chaotic arrest? Probably not. First, it is questionable whether people gathering around the scene of an arrest constitutes a riot. Second, it is uncertain whether there was probable cause to believe that the man taking pictures caused the people to gather around. It seems more likely that the arrest itself attracted the attention of passersby. A court would likely find that the officers did not have probable cause to believe that the man taking pictures of the arrest provoked a riot.[223]

In some cases where police officers arrest photographers pursuant to broadly worded statutes,[224] factual disputes and problems of proof may arise. For example, suppose Person A is taking pictures on a public sidewalk thirty feet away from a car that is being searched by police for drugs and weapons. The police officers demand that Person A stop taking pictures or risk being arrested. Person A is convinced he is not breaking any laws and continues to take pictures. The police officers arrest him for obstruction of justice. The obstruction of justice statute prohibits interfering with a police investigation. The statute defines "interfering" as coming within fifteen feet of an investigation or arrest for purposes of photographing the scene. Determining whether police had probable cause to believe Person A was obstructing justice will likely come down to Person A's word against the word of the police officers. Given these facts, a court likely will show deference to the police officers.[225] It is possible that Person

220. *See McClain*, 444 F.3d at 562.

221. *Id.*

222. *Id.*

223. *See Harris*, 806 A.2d at 127–28.

224. *See* Powell, *supra* note 151, at 19; *see also* Burton, *supra* note 87, at 19–21.

225. *See* Angela J. Davis, *Prosecution and Race: The Power and Privilege of Discretion*, 67 FORDHAM L. REV. 13, 27–28 (1998) ("Despite the requirement that a police

A could have an expert view the pictures and testify that they were taken from a distance of much more than fifteen feet. However, this may be difficult considering the zoom function on most cameras. It would also be extremely costly, and probably not worth the amount Person A would recover in damages.

In sum, a § 1983 plaintiff has a stronger chance of establishing a Fourth, rather than a First Amendment violation. However, factual disputes and difficulties in establishing proof will arise. A plaintiff will first have to show that he was seized.[226] The plaintiff will then need to show that no reasonable suspicion existed for the seizure.[227] If the detention was not pursuant to any statute or regulation, proving the absence of reasonable suspicion will not be problematic. On the other hand, if the detention was pursuant to something more than, for example, "national security purposes," showing no reasonable suspicion existed will be more difficult. Reasonable suspicion is a low standard and courts often show deference to police officers.[228]

2. Qualified Immunity

Establishing a constitutional violation is only the first of a few hurdles one must clear to ultimately succeed in a § 1983 suit. The next hurdle for a plaintiff is qualified immunity, which is enjoyed by both state and federal officials. Federal officials are not liable for violations of constitutional rights committed in "good faith."[229] In *Smith v. Kenny*, the federal district court in New Mexico stated that "courts recognize the affirmative defense of qualified immunity, which protects 'all but the plainly incompetent or those who knowingly violate the law.'"[230] The *Smith* court set out a two-part test for qualified immunity: "Once a defendant asserts the affirmative defense of qualified immunity, the burden then shifts to the plaintiff to establish (1) a violation of a constitutional or federal statutory right by the defendant, and (2) that the constitutional right allegedly violated was

---

officer's decision to stop a suspect must be based on an articulable suspicion, the Supreme Court has shown increasing deference to the judgment of police officers in its interpretation of this requirement. The practical effect of this deference is the assimilation of police officers' subjective beliefs, biases, hunches, and prejudices into law.") (citations omitted).

226. *See* Brown v. City of Oneonta, 221 F.3d 329, 340 (2d Cir. 2000).

227. *See* United States v. Gray, 213 F.3d 998, 1000 (8th Cir. 2000) (noting that both a search and seizure are "constitutionally reasonable" when "based upon reasonable suspicion that criminal activity is afoot."); *Jones*, 745 A.2d at 861 (citing *Terry*, 392 U.S. at 21) ("[A] police officer may detain an individual for investigatory purposes for a limited scope and duration, but only if such detention is supported by a reasonable and articulable suspicion of criminal activity.").

228. *See* Davis, *supra* note 225, at 27–28.

229. *See* Courson v. McMillian, 939 F.2d 1479, 1487 (11th Cir. 1991) (citing Rich v. Dollar, 841 F.2d 1558, 1563 (11th Cir. 1988)).

230. Smith v. Kenny, 678 F. Supp. 2d 1093, 1105 (D.N.M. 2009).

clearly established at the time of the violation."[231] The previous section addressed the many difficulties that arise in establishing a constitutional violation;[232] this section will focus on the second prong of the analysis.

The inquiry as to whether a plaintiff's right was "clearly established" is one that focuses on factual correspondence between the alleged unlawful actions and case law.[233] Officers lose qualified immunity when the unlawfulness of their actions was apparent[234] or when there was no legitimate question as to the unlawfulness of the conduct. [235] Officials may lose their immunity even if "the very action in question" had not been declared unlawful.[236] Courts are split on what type of authority may render a right clearly established. Some courts have stated that only decisions by the Supreme Court, a court of appeals, or a state's highest court may indicate that a right is clearly established.[237] Additionally, some courts have held that a constitutional provision that requires a balancing of interests is generally not a clearly established right.[238] The Ninth Circuit has held that it is the "[p]laintiff's burden . . . to identify the universe of statutory or decisional law from which the court can determine whether the right allegedly violated was clearly established."[239]

A §1983 plaintiff's success in showing that a right is clearly established at the time of a violation depends on which right the plaintiff claims and the manner in which the plaintiff frames this right. It will be easier to demonstrate that a plaintiff's right was clearly established under the Fourth Amendment than under First Amendment. Freedom of the press is "clearly established," but the extent to which the First Amendment protects newsgathering activities is more ambiguous.[240] Most courts agree that newsgathering qualifies for First Amendment protection because "[w]ithout some protection seeking out the news, freedom of the press could be eviscerated."[241] However, the Supreme Court has also stated that, "[t]he [First] [A]mendment does not reach so far as to override the interest of the public in ensuring that neither the reporter nor the source is invading the rights of other citizens through reprehensible conduct forbidden to all other

231. *Id.* at 1106.

232. *See supra* Section IV.C.1.

233. *See* Benson v. Allphin, 786 F.2d 268, 276 (7th Cir. 1986).

234. Anderson v. Creighton, 483 U.S. 635, 639–40 (1987).

235. Mitchell v. Forsyth, 472 U.S. 511, 535 n.12 (1985) (noting when there is a legitimate question as to the legality of an official's actions at the time the official engages in the actions, the conduct "cannot be said [to] violate[] clearly established law").

236. *See Anderson*, 483 U.S. at 639–640 ("This is not to say that an official action is protected by qualified immunity unless the very action in question has previously been held unlawful . . . .").

237. Courson v. McMillian, 939 F.2d 1479, 1498 n.32 (11th Cir. 1991).

238. *Benson*, 786 F.2d at 276.

239. Elder v. Holloway, 975 F.2d 1388, 1392 (9th Cir. 1991).

240. Ugland, *supra* note 184, at 121.

241. Branzburg v. Hayes, 408 U.S. 655, 693 (1972).

persons."[242] Professor Erik Ugland writes that there is an "[a]bsence of uniformity [among courts] regarding whether the First Amendment provides merely a negative barrier against government intrusions or also provides a set of affirmative rights—rights of access to places and information . . . ."[243] He notes that many courts "uniformly shape rights by a negative construction."[244]

Suppose Person A is photographing police who are trying to control a political protest in a city park. Police demand that he stop, but he continues. Person A is arrested for disorderly conduct and obstruction of justice. He will need to demonstrate that his right to photograph political protests for news purposes is clearly established. This will be hard considering most courts have given only passive support to any affirmative right to gather the news.[245] Under the Seventh Circuit's interpretation, since the right to gather the news requires a balancing of interests, it is not clearly established for the purposes of qualified immunity.[246] Courts have made it clear that any protections afforded to the newsgathering process do not provide a shield from liability under criminal statutes.[247] Consequently, even if a court were to recognize newsgathering as a clearly established right, it may find that this right stops at the point where Person A's actions constitute disorderly conduct or obstruction of justice.

Showing that the right to freedom of photographic expression is clearly established is even more difficult. The Supreme Court has held that "[p]hotography, painting, and other two-dimensional forms of artistic reproduction . . . are plainly expressive activities that ordinarily qualify for First Amendment protection."[248] The Supreme Court has also found that "works which, taken as a whole, possess serious artistic value are protected by the First Amendment."[249] However, other courts have held that "images . . . must communicate some idea in order to be protected under the First Amendment."[250] Although it is clearly established that photographs may fall within the scope of First Amendment protection, it is not clearly established that the *act* of taking the photograph that may later be used to express ideas is a right secured by the First Amendment.[251] As a whole, courts have been

242. *Id.* at 691–92.

243. Ugland, *supra* note 184, at 139.

244. *Id.* at 143.

245. *Id.* at 121.

246. Benson v. Allphin, 786 F.2d 268, 276 (7th Cir. 1986).

247. For example, the Sixth Circuit has stated that "The [F]irst [A]mendment of course does not immunize wrongful behavior simply because it is undertaken in the name of newsgathering." Boddie v. Am. Broad. Co., 881 F.2d 267, 271 (6th 1989).

248. Massachusetts v. Oakes, 491 U.S. 576, 591 (1989).

249. *See* Miller v. California, 413 U.S. 15, 24 (1973).

250. Montefusco v. Nassau Cnty., 39 F. Supp. 2d 231, 241 (E.D.N.Y. 1999) (citing Berry v. New York, 97 F.3d 689, 695 (2d Cir. 1996)).

251. *See id.* at 242 n.7.

reluctant to extend First Amendment protection to plaintiffs who have claimed that they do not plan to use photographs they have taken for expressive and transformative purposes.[252]

In sum, arguing that First Amendment rights within the newsgathering and pre-expressive photography contexts are clearly established will be difficult. Although courts have extended some First Amendment protection to newsgathering and photographic expression, these protections are loaded with caveats and qualifications.[253] In most cases, a balancing of interests is involved, and many courts are hesitant to consider rights clearly established that require case-by-case balancing.[254]

A § 1983 plaintiff has a better chance of convincing a court that rights protected under the Fourth Amendment are clearly established, as Fourth Amendment protection against unreasonable searches and seizures is clearly established law.[255] Thus, the issue in establishing the second prong of the qualified immunity test generally will be "whether or not the officers made a reasonable mistake as to what the law requires."[256] In *Hudson v. Felder*, the United States District Court for the Eastern District of Kentucky held that "[i]f a police officer . . . arrests a citizen where probable cause is so clearly absent that the officer sheds his or her qualified immunity, the officer may be held accountable under § 1983 for the wrongful conduct."[257] The Supreme Court illuminates when probable cause might be found: "Probable cause exists where 'the facts and circumstances within . . . [the agent's] knowledge and of which they had reasonably trustworthy information [are] sufficient in themselves to warrant [one] of reasonable caution in the belief that an offense has been or is being committed."[258] Reasonableness is evaluated from the perspective of a government actor at the scene, not with the benefit of hindsight.[259]

In order for an officer to have a reasonable "belief that an offense has been or is being committed," there must be an underlying offense.[260] If the

---

252. *See id.*

253. *See id.* at 242 (citing *Hurley v. Irish-Am. Gay, Lesbian and Bisexual Group of Boston*, 515 U.S. 557, 569, (1995); *See also* Ugland *supra* note 184, at 121 (noting that courts are not at a consensus as to the protection provided to newsgathering under the First Amendment).

254. *See* Benson v. Allphin, 786 F.2d 268, 276 (7th Cir. 1986).

255. Gale v. Storti, 608 F. Supp. 2d 629, 633 (E.D. Pa. 2009). *See also* Herren v. Bowyer, 850 F.2d 1543, 1547 (11th Cir. 1988) ("The law is 'clearly established' that an arrest without a warrant or probable cause to believe a crime has been committed violates the [F]ourth [A]mendment . . . .").

256. *Gale*, 608 F. Supp. 2d at 634.

257. Hutson v. Felder, No. 5:07-183-JMH 2008 WL 4186893, at *3 (E.D. Ky. Sept. 10, 2008).

258. Brinegar v. United States, 338 U.S. 160, 175–76 (1949) (quoting Carroll v. United States, 267 U.S. 132, 162 (1925)).

259. Graham v. Connor, 490 U.S. 386, 396 (1989).

260. *Brinegar*, 338 U.S. at 175–76.

arrest is not made pursuant to an offense defined by statute, law, or regulation, then an officer cannot have a reasonable belief that an offense has been or is being committed and probable cause would not exist.[261] Thus, if Person A, discussed in the hypothetical above, is arrested pursuant to a vague assertion of national security for taking pictures of bridges or buildings, a court may find that "probable cause is so clearly absent that the officer [has] shed[] his or her qualified immunity."[262]

The right to be free from unlawful detention is "clearly established" for the purposes of clearing the qualified immunity hurdle.[263] In most of the incidents examined in this article, however, people taking pictures of bridges, buildings, national laboratories, etc., are never ultimately arrested.[264] They are detained in many instances, which requires a "reasonable suspicion that the person has been, is, or is about to be engaged in criminal activity."[265] As the federal district court in Arizona held, "Reasonable suspicion to support an investigatory stop exists when an officer is aware of specific, articulable facts which, when considered with objective and reasonable inferences, form a basis for particularized suspicion."[266]

In the Portland incident described earlier (involving the high school girls who were held and questioned for photographing oil tankers near bridges),[267] the suspected criminal activity in this case was presumably potential terrorist activity.[268] The only articulable fact an officer could base reasonable suspicion was the fact that it appeared as if the girls were taking pictures of a bridge.[269] Courts have held that "[a]n individual's fundamental Fourth Amendment right to be free from 'unreasonable searches and seizures' does not dissipate merely because of generalized, unsubstantiated suspicions of terrorist activity."[270] Therefore, in the Portland case, the plaintiffs would likely be able to show a violation of a clearly established right.

Showing a violation of a clearly established Fourth Amendment right is much harder when a plaintiff is charged under a broadly worded criminal statute.[271] In such a case, there will be more factual disputes over what

---

261. *See Herren*, 850 F.2d at 1545–46.

262. *Id.*

263. *See* Scheier v. City of Snohomish, No. C07-1925-JCC, 2008 WL 4812336, at *6 (W.D. Wash. Nov. 4, 2009).

264. *See supra* text accompanying notes 25–33 & 35–54.

265. Michigan v. Summers, 452 U.S. 692, 698 n.7 (1981).

266. Ortega Melendres v. Arpaio, 598 F. Supp. 2d 1025, 1026 (D. Ariz. 2009).

267. *See supra* text accompanying notes 11–16.

268. *Id.*

269. *See supra* text accompanying notes 14–15.

270. Scheier v. City of Snohomish, No. C07-1925-JCC, 2008 WL 4812336, at *11 (W.D. Wash. Nov. 4, 2009).

271. *See supra* note 151 and accompanying text.

actually happened. Moreover, the breadth of such statutes often makes it easy to build a case against a plaintiff since many activities can reasonably be construed as being criminal under sufficiently broad language.[272] For purposes of qualified immunity, the Supreme Court has held that even when a plaintiff perceives an officer's conduct to be malicious, immunity may attach if the officer's conduct is found to be *objectively* reasonable.[273] In cases where an officer "reasonably but mistakenly" believed that probable cause existed to effect the arrest or that certain exceptions applied to justify an arrest absent probable cause, the officer generally will be relieved from liability on the basis of qualified immunity.[274] Thus, it is easy to imagine how vague statutory language may unintentionally (or perhaps intentionally) provide law enforcement with the ability to claim for their conduct objective reasonableness and hence curtail civil liberties under the guise of "national security."

If a court were to determine that an officer "reasonably but mistakenly" believed that a journalist's conduct constituted disorderly conduct or obstruction of justice, the officer will not be subject to liability.[275] State statutes prohibiting disorderly conduct and obstruction of justice are generally broadly written[276] and consequently cover a wide range of conduct. A Michigan city ordinance makes it unlawful to "[a]ssault, obstruct, resist, hinder, or oppose any member of the police force . . . ."[277] The Sixth Circuit upheld qualified immunity where an officer arrested a person pursuant to this ordinance for interrupting him during questioning of a third party.[278] This court also upheld qualified immunity under circumstances where an officer arrested a person for refusing to provide identification.[279] In *Washpon v. Parr*, the United States District Court for the Southern District of New York upheld qualified immunity where officers arrested and charged the plaintiff with disorderly conduct for allegedly disobeying officers' orders to leave a courthouse.[280] With respect to whether the officers had probable cause to arrest the plaintiff, the *Washpon* Court found that "reasonably competent police officers could disagree,"[281] and thus the officer was entitled to qualified immunity.[282]

---

272. *See e.g.*, incidents described *supra* notes 17–24, 34–39, 55–63.

273. *See* Malley v. Briggs, 475 U.S. 335, 341 (1986); *see* Harlow v. Fitzgerald, 457 U.S. 800, 807 (1982) (holding that an allegation of malice is not sufficient to defeat immunity if the defendant acted in an objectively reasonable manner).

274. Anderson v. Creighton, 483 U.S. 635, 641 (1987).

275. *See id.*

276. *See* Burton, *supra* note 87, at 19.

277. Risbridger v. Connelly, 275 F.3d 565, 568 (6th Cir. 2002) (quoting EAST LANSING, MICHIGAN, Code, Title IX, Ch. 108, § 9.102(19)).

278. *See* King v. Ambs, 519 F.3d 607, 608 (6th Cir. 2008).

279. *See Risbridger*, 275 F.3d at 567.

280. Washpon v. Parr, 561 F. Supp. 2d 394, 395 (S.D.N.Y. 2008).

281. *Id.* at 403.

282. *Id.* at 404.

Other courts have upheld qualified immunity where no probable cause existed for the crime the plaintiff was arrested for but existed for some other criminal charge. In *Ware v. James City County, Virginia*, for example, the United States District Court for the Eastern District of Virginia found that the "[defendant's] initial reason for making the arrest need not be the criminal offense that ultimately is supported by probable cause from the known facts."[283] The aforementioned cases demonstrate the extent to which courts show deference to police officers when considering whether or not to uphold qualified immunity.

The obstacle of qualified immunity can be overcome when police officers detain or arrest persons for some vague assertion of "national security concerns" or unsubstantiated suspicions of "terrorist activity."[284] Additionally, a plaintiff may be able to overcome qualified immunity if he or she has been charged under a broad criminal statute, and it is obvious that officers were acting outside of accepted practice or perhaps had nefarious motives for the arrest (e.g., preventing a photographer from publishing pictures of police misconduct). Otherwise, overcoming the bar of qualified immunity is difficult. Although Fourth Amendment rights are "clearly established," police officers are not liable unless they violate such rights unreasonably.[285] Doing so is difficult, considering that most of the broadly worded statutes used to charge journalists and photographers are all encompassing and criminalize a broad spectrum of conduct.[286] In addition, courts tend to be highly deferential to police officers so as to prevent a judgment against law enforcement officials conducting their jobs as we would expect.[287] Even assuming a plaintiff does make it past the nearly impenetrable barrier of qualified immunity, he or she must still prove damages, which can be difficult.

3. Proving Damages

The U.S. District Court for the District of Columbia explained that "[t]he purpose of a damage award under § 1983 is 'to compensate persons for injuries that are caused by the deprivation of constitutional rights.'"[288] Indeed, "[c]ompensatory damages may include out-of-pocket loss,

283. Ware v. James City Cnty., Virginia, 652 F. Supp. 2d 693 (E.D. Va. 2009).

284. Scheier v. City of Snohomish, No. C07-1925-JCC, 2008 WL 4812336, at *10 (W.D. Wash. Nov. 4, 2008).

285. *Id.* at *6.

286. *See* Burton, *supra* note 87, at 19; *see also* incidents cited *supra* notes 17–24, 34–39, 55–63 (providing examples of broadly worded criminal statutes used to bring charges against photographers).

287. Teressa E. Ravenell, *Cause and Conviction: The Role of Causation in § 1983 Wrongful Conviction Claims*, 81 TEMP. L. REV. 689, 709 (2008).

288. Elkins v. District of Columbia, 610 F. Supp. 2d 52, 60 (D.D.C. 2009) (quoting Carey v. Piphus, 435 U.S. 247, 254 (1978)).

impairment of reputation, humiliation, and mental anguish and suffering."[289] In addition, the Ohio Court of Appeals has held that "[a] prevailing party [in a § 1983 action] should be allowed attorney fees unless 'special circumstances' would render awarding fees unjust."[290] Professor Teressa Ravenell provides a helpful background on the Supreme Court's shift in § 1983 cases:

> In *Carey v. Piphus*, the U.S. Supreme Court held that a plaintiff alleging a deprivation of a procedural due process right must show that the deprivation resulted in an actual injury in order to receive more than nominal monetary damages.[291] Shortly thereafter, the Court extended its holding to all Section 1983 cases for money damages, regardless of the underlying constitutional deprivation.[292] Thus, to receive monetary damages, all Section 1983 plaintiffs must prove that they suffered an actual injury.[293]

As alluded to above, proving damages can present legal hurdles in the context of these photography cases. Suppose Person A tours the United States for three weeks, accumulating hundreds of pictures on his memory card. At his last stop, an officer detains him and deletes all of his pictures for suspected terrorist activity. He successfully brings a § 1983 claim. Person A may never visit any of the places on his trip again, and he considers the deleted pictures to be priceless. The nature of his losses renders full compensation impossible, for the sentimental value of the photographs may be irreplaceable. Likewise, in the case of the girls taking pictures of a bridge in Portland,[294] does existing tort law provide a remedy for being placed on an FBI terrorist watch list?

The first problem Person A will face is the nearly impossible task of convincing a judge or jury of an "evidentiary link between the defendant's breach and [Person A's] injury."[295] Some courts apply a tort-based approach to causation in § 1983 cases, asking "whether the defendant should have foreseen that his conduct would result in the plaintiff's injury."[296] In applying this approach, a court may find that the officer who arrested Person A could not have reasonably foreseen the loss experienced by Person A. The court would probably agree with the defendant's argument that the officer reasonably would not have known that Person A was a journalist and intended to publish the photos (although it seems

289. *Id.* at 59 (citing Memphis Cmty. Sch. Dist. V. Stachura, 477 U.S. 299, 307 (1986)).

290. Thomas v. City of Cleveland, 892 N.E.2d 454, 458 (Ohio App. 8 Dist. 2008).

291. Carey v. Piphus, 435 U.S. 247, 264 (1978).

292. Memphis Cmty. Sch. Dist. v. Stachura, 477 U.S. 299, 309–10 (1986).

293. Ravenell, *supra* note 287, at 709.

294. *See supra* text accompanying notes 11–16.

295. Ravenell, *supra* note 287, at 714.

296. *Id.* at 722.

obvious that this would be the precise reason the officer deleted the pictures). Therefore, a court would likely agree with the defendant's categorization of Person A's alleged damages as speculative. As Professor Ravenell notes, "Courts have used [damages] causation to limit liability in § 1983 wrongful conviction claims."[297]

All said, § 1983 plaintiffs within the context of the War on Photography are not likely to receive full compensation for their injuries.[298] A plaintiff may be able to collect for mental anguish and humiliation resulting from being arrested or detained, but showing any further harm will be difficult. While the true value of a journalist's pictures is not fully realized until after they are published, the constitutional violation due to their destruction occurs much before then.[299] Courts are likely to discard any arguments regarding the potential uses of the photographs as speculative and attenuated and perhaps unforeseen—thereby dealing a final, crucial blow to establishing legal causation.

4. § 1983 Summary

In most cases, § 1983 will not provide an adequate remedy for a plaintiff suffering a First or Fourth Amendment violation within the context of taking photographs in public places. First, establishing a constitutional violation will be difficult. Courts have given only passive support to § 1983 claims involving photography rights under the First Amendment.[300] Courts have been even more hesitant to extend any meaningful First Amendment protections to the newsgathering process, and where they have, the protections are heavily qualified.[301] A plaintiff has a much better chance of establishing a Fourth Amendment violation, but doing so may involve factual disputes and difficulty in providing the necessary proof.[302] In many cases, it will amount to the officer's word against the plaintiff's, and courts are generally deferential to officers.[303]

Second, qualified immunity represents a major obstacle to a § 1983 plaintiff.[304] Police officers are armed with broadly written criminal statutes that can be interpreted to prohibit a wide range of conduct,[305] in addition to

297. Ravenell, *supra* note 287, at 693.

298. *See supra* text accompanying notes 295–97.

299. *But see, e.g.*, Montefusco, 39 F. Supp. 2d 231, 242 n.7 (asserting that it is not clearly established whether the act of taking a photograph is protected by the First Amendment).

300. *See Montefusco*, 39 F. Supp. 2d at 242 n.7; *see also supra* notes 180–83, 185–86 and accompanying text.

301. *See supra* notes 180–191 and accompanying text.

302. *See supra* notes 192–204 and accompanying text.

303. *See supra* note 225 and accompanying text.

304. *See supra* note 229 and accompanying text.

305. *See, e.g.*, incidents cited *supra* notes 17–24, 34–39, 55–63.

the deference of courts.[306] Even if an officer commits a Fourth Amendment violation, as long as the officer was not completely incompetent in doing so,[307] a court will probably find that he acted reasonably under the circumstances and is thus entitled to qualified immunity.[308]

Finally, assuming a plaintiff successfully clears these § 1983 obstacles, she still must prove damages. Professor Ravenell has noted that courts use damages causation to limit the scope of liability of state officials under § 1983.[309] The nature of photography and journalism is not conducive to proving damages.[310] First of all, it is difficult to put a dollar value on photographs that are worth different things to different people. One may consider photographs from a three-week vacation to be worth a great deal. A judge or juror may not fully appreciate this, and compensation will reflect that. Secondly, the value of photographs is realized upon publication; however, courts may consider the time in between the violation and publication and the process of publication itself as enough to render damages speculative.[311] Courts may also use causation to limit damages by finding that a defendant could not have foreseen that the plaintiff was (1) a journalist (2) who intended to publish the photographs, and (3) would be personally and professionally damaged by having his or her pictures deleted.[312] Within the context of violations occurring while photographing police misconduct, this seems naive at best. Arguably, a violation itself occurs because of the officer's realization that a journalist is snapping pictures that will be published and expose the officer's misconduct. So to hold in this context that the defendant could not have foreseen the damages is illogical.

Because proving damages involving photographs is inherently difficult, a plaintiff may collect only nominal damages or damages for humiliation and embarrassment resulting from the arrest itself.[313] Nominal damages are significant since they are generally accompanied with attorney fees,[314] and damages for humiliation and embarrassment will provide some relief for a plaintiff. However, jumping through all of the hoops required to succeed in a § 1983 action is a lengthy, arduous, and costly process. When litigation is commenced, a plaintiff has no guarantee that the days, weeks, and sometimes years her attorney spends on the case will be paid for by someone other than herself. Even if attorney's fees are provided, litigation

306. *See, e.g.*, *supra* notes 225, 228, 278, 280–83 (providing examples of courts showing deference to police officers).

307. *See supra* note 230 and accompanying text.

308. *See supra* notes 255–59 and accompanying text.

309. *See* Ravenell, *supra* note 287, at 733.

310. *See supra* text accompanying notes 295–97.

311. *See supra* text accompanying notes 288–90.

312. *See supra* text accompanying notes 295–97.

313. *See supra* note 289 and accompanying text.

314. *See supra* note 290 and accompanying text.

is time-consuming. Given the burdens of litigation, § 1983 does not provide a meaningful remedy to most plaintiffs suffering constitutional violations for taking pictures in public places.

D. Bivens *Action*

In *Bivens v. Six Unknown Named Agents of Federal Bureau of Narcotics*, the Supreme Court established a federal common law cause of action for damages caused by a federal agent acting "under color of his authority" in violation of a claimant's constitutional right.[315] If a federal official violates a person's constitutional rights, a plaintiff cannot recover under § 1983, and must instead bring a *Bivens* action. "A *Bivens* action is a nonstatutory counterpart of a suit brought pursuant to 42 U.S.C. Section 1983, and is aimed at federal rather than state officials."[316] *Bivens* actions apply in limited settings, generally to violations of the Fourth and Fifth Amendments.[317] Unfortunately, the Supreme Court has been reluctant to extend *Bivens* liability to violations of the First Amendment, precisely those that would presumably apply to photography cases involving journalists, bloggers, etc.[318]

---

315. Bivens v. Six Unknown Named Agents of Fed. Bureau of Narcotics, 403 U.S. 388, 389 (1971). Factors that strongly suggest that a police encounter has become a seizure include:

> [T]he threatening presence of several of officers; the display of a weapon; physical touching of the person by the officer; language or tone indicating that compliance with the officer was compulsory; prolonged retention of a person's personal effects . . . and a request by the officer to accompany him to the police station or a police room.

Brown v. City of Oneonta, 221 F.3d 329, 340 (2d Cir. 2000) (quoting United States v. Hooper, 935 F.2d 484, 491 (2d Cir. 1991)). This remedy stands in contrast to most of the others examined herein in that it applies specifically to federal agents; the other scenarios and incidents explored here have involved state law enforcement or private security guards.

316. Mahoney v. Nat'l Org. for Women, 681 F.Supp. 129, 132 (D.Conn. 1987).

317. Davis v. Passman, 442 U.S. 228, 229–30 (1979).

318. Ashcroft v. Iqbal, 129 S. Ct. 1937, 1939 (2009). The *Ashcroft* Court states:

> For while we have allowed a *Bivens* action to redress a violation of the equal protection component of the Due Process Clause of the Fifth Amendment, we have not found an implied damages remedy under the Free Exercise Clause. Indeed, we have declined to extend *Bivens* to a claim sounding in the First Amendment. Petitioners do not press this argument, however, so we assume, without deciding that respondent's First Amendment claim is actionable under *Bivens*.

Id. (internal citations omitted).

The requirements for successfully bringing a *Bivens* claim are stringent.[319] A plaintiff must first allege specific facts sufficient to support a violation of a right secured by the Constitution,[320] and the failure to allege such a constitutional violation is "fatal to their case."[321] A plaintiff must next show that his or her constitutional right was violated by an agent of the United States acting under color of law.[322] When "a constitutionally recognized interest is adversely affected by the actions of federal employees," the Court applies a two-pronged analysis: (1) is there an alternative judicial process that can '"protect the interest" which is "convincing" enough for the Court to refrain from providing a new remedy; or (2) if there is no convincing alternative process, are there any "special factors counseling hesitation before authorizing a new kind of federal litigation."[323] If the answer to the first prong is yes, then a new remedy will not be created.[324] If the answer is no, the court will consider "special factors"[325] such as the adequacy of alternative remedies,[326] difficulty in defining legitimate action by government actors,[327] the importance of protecting the constitutional interest,[328] the demand and cost on the judicial system from creating a mass of new litigation in the area,[329] the difficulty in defining a broader doctrine,[330] and the ability of Congress to legislate a remedy.[331] Through a successful *Bivens* claim, a plaintiff may collect money damages from agents of the United States in their individual capacities.[332]

The *Bivens* remedy is a wholly inadequate solution to the growing problem of photography rights violations. The remainder of this subsection will discuss both the general obstacles a *Bivens* plaintiff faces and also the specific obstacles present within the context of photography.

---

319. Duxbury Trucking, Inc. v. Mass. Highway Dep't, No. 04cv12118-NG, 2009 WL 1258998, at *2 (D. Mass. Apr. 29, 2009).

320. *Id.* (citing *Mahoney*, 681 F. Supp. at 132).

321. Mattox v. City of Forest Park, 183 F.3d 515, 520 (6th Cir. 1999).

322. *Id.*

323. Wilkie v. Robbins, 55 U.S. 537, 550 (2007).

324. *See id.*

325. *See id.*

326. *See id.* at 555.

327. *See id.*

328. *See id.* at 577.

329. *See id.*

330. *See id.*

331. *See id.* at 562.

332. *See id.* at 575.

1. General Obstacles to *Bivens* Recovery

Between 1971 and 1989, twelve thousand *Bivens* suits were filed and only thirty resulted in judgments on behalf of the plaintiffs.[333] A number were reversed on appeal and only four judgments were actually paid by the individual federal defendant.[334] According to Perry M. Rosen, a former trial attorney for the U.S. Department of Justice specializing in *Bivens* cases, a number of factors explain these statistics.

First, plaintiffs in *Bivens* suits are procedurally disadvantaged.[335] A "*Bivens* plaintiff must plead the alleged constitutional tort with greater specificity than other claims."[336] Additionally, courts have "construed jurisdiction, venue, and other preliminary issues in *Bivens* suits so as to favor the individual government defendant."[337] Second, judges and juries are extremely reluctant to render judgments in favor of the plaintiff when they know that such a judgment will result in thousands of dollars out of pocket for a federal employee.[338] Judges and juries are aware that federal officials often have limited resources and do not want to pin a sizable judgment on an official who may have been simply trying to do his job.[339] Third, juries are less likely to side with plaintiffs in *Bivens* actions because it is "more difficult to 'see' the injury from a constitutional tort . . . then [sic] the injury from a common-law tort involving personal injury."[340] Fourth, federal officials have qualified immunity and are not liable for violations of constitutional rights committed in "good faith."[341] Fifth, even if a plaintiff receives a judgment in its favor, it is far from certain that the federal agent will have funds to compensate the plaintiff.[342] The "deep pockets" of the federal government are not available to *Bivens* plaintiffs.[343]

2. Obstacles to *Bivens* Plaintiffs within the Context of Photography Rights

Alleging specific facts sufficient to support a constitutional violation will be difficult for a number of reasons. As stated earlier, courts have been somewhat reluctant to extend *Bivens* liability to First Amendment

---

333. Perry M. Rosen, *The* Bivens *Constitutional Tort: An Unfulfilled Promise*, 67 N.C. L. Rev. 337, 343 (1989).
334. *Id.* at 343–44.
335. *Id.* at 345.
336. *Id.*
337. *Id.*
338. *Id.* at 347.
339. *Id.*
340. *Id.*
341. *Id.* at 348–49.
342. *Id.* at 347.
343. *Id.*

violations.[344] It is unclear how receptive a court would be to the claim that federal officials violated one's right to photographic expression or freedom of the press within the newsgathering context. If a plaintiff makes it past this initial hurdle, he or she still must show that there is no judicial process that can protect his or her interest that is convincing enough for the court to refrain from providing a new remedy.[345] On this point, the Supreme Court has said that "[s]o long as the plaintiff [has] an avenue for some redress, bedrock principles of separation of powers foreclose[] judicial imposition of a new substantive liability."[346] Since Congress has created no statutory scheme addressing First Amendment violations in the context of photography, a court would likely find that no convincing remedy exists.

Congress created a remedy for certain torts committed by federal employees in the Federal Tort Claims Act.[347] The Act gives the district courts jurisdiction over the loss of property "caused by the negligent or wrongful act or omission of any employee of the Government while acting within the scope of his office or employment . . . ."[348] Current law also provides that "[t]he United States shall be liable . . . in the same manner and to the same extent as a private individual under like circumstances . . . ."[349] But the Court has found it to be "'crystal clear' that Congress intended the FTCA and *Bivens* to serve as 'parallel' and 'complementary' sources of liability."[350] Courts are unlikely to preclude a plaintiff's *Bivens* claim solely based on the availability of a claim under the FTCA.[351]

Next, a plaintiff must maneuver around a number of "special factors" that may counsel hesitation before a court will authorize a new kind of federal litigation.[352] Recently, scholars have noted that courts have broadly interpreted the "special factors" to narrow the scope of *Bivens*.[353] Natalie Banta argues that the *Bivens* test has become "more legislative than judicial in nature, because the Court can now make policy decisions as to whether or not to apply the remedy instead of looking solely to the remedies available and assessing whether they are adequate."[354] The special factors established in *Wilkie v. Robbins* are expansive, leaving courts with virtually

344. Ashcroft v. Iqbal, 129 S. Ct. 1937, 1947–48 (2009).

345. Rosen, *supra* note 333, at 357.

346. Corr. Servs. Corp. v. Malesko, 534 U.S. 61, 69 (2001).

347. *See* Federal Tort Claims Act, 28 U.S.C. § 1346(b) (2006).

348. *Id.*

349. 28 U.S.C. § 2674 (2006).

350. *Corr. Servs. Corp.*, 534 U.S. at 68.

351. *Cf.* Rosen, *supra* note 333, at 358 (noting other circumstances in which the Court has allowed a plaintiff to simultaneously maintain her actions under the FTCA and the Eighth Amendment).

352. *Id.* at 359.

353. Natalie Banta, *Death by a Thousand Cuts or Hard Bargaining?: How the Court's Indecision in* Wilkie v. Robbins *Improperly Eviscerates the Bivens Action*, 23 BYU J. PUB. L. 119, 135 (2008).

354. *Id.*

unlimited discretion in fashioning reasons to reject *Bivens* claims.[355] A court could easily reject a photographer's *Bivens* claim based on broad, inarticulate concerns such as avoiding a flood of litigation or difficulty in judicially defining a broad doctrine.[356]

Assuming a plaintiff can prove facts sufficient to support the assertion of a constitutional violation, demonstrate that no other adequate remedy exists, and navigate through the maze of special factors, he or she must next survive the doctrine of qualified immunity, which in most *Bivens* cases is fatal.[357] As stated earlier, federal employees possess qualified immunity and are not held liable for violations of constitutional rights committed in good faith.[358] A state actor is not entitled to qualified immunity if: (1) the plaintiff's allegations, assuming they are true, establish a constitutional violation, (2) the constitutional right at issue was clearly established at the time of the putative violation, and (3) a reasonable officer, situated similarly to defendant, would have understood the challenged act or omission to contravene the discerned constitutional right.[359] Since qualified immunity in *Bivens* actions is essentially the same doctrine applied in § 1983 cases, and since the previous section explains the ways in which qualified immunity drastically reduces a plaintiff's chances of securing relief, no further discussion will be provided in this section.

3. *Bivens* Summary

In most cases, a *Bivens* action will not provide a plaintiff with an adequate remedy in the context of photography rights. The two most obvious constitutional provisions applicable to photography rights in public places include the First Amendment freedom of the press and freedom of expression. Although some courts have extended the scope of *Bivens* to cover First Amendment violations, no court has recognized a *Bivens* action for violation of freedom of the press due to interference with the "newsgathering" process.[360] Nor has a court extended *Bivens* to a First Amendment violation of photographic expression. Even if a plaintiff successfully alleges a constitutional violation that cannot be remedied by an alternative judicial process, a court may still use one of a number of special factors to avoid extending the scope of *Bivens* liability.[361] If a court does recognize the action, qualified immunity presents a nearly impenetrable

355. Wilkie v. Robbins, 551 U.S. 537, 555, 561 (2007).

356. *See id.* at 555, 561.

357. *See* Rosen, *supra* note 333, at 348.

358. *See id.* at 348–49.

359. Savard v. Rhode Island, 338 F.3d 23, 27 (1st Cir. 2003).

360. *See generally* Spagnola v. Mathis 809 F.2d 16, 32 (D.C. Cir. 1988) (reversing dismissal of First Amendment *Bivens* claim).

361. *See* Rosen, *supra* note 333, at 359.

barrier that judges and juries are more than willing to apply.[362] Even a plaintiff who successfully brings a *Bivens* suit may not be compensated fully.[363] Federal employees do not have the "deep pockets" of the federal government and may not be able to afford the judgment against them.[364] Most important of all, attorneys' fees are not available in *Bivens* actions.[365] A *Bivens* judgment would have to be quite large to cover both the attorneys' fees (which will likely be substantial) and any damages the plaintiff suffered. As explained within the context of § 1983, damages associated with photography rights are inherently difficult to prove and courts use causation to limit the damages for which *Bivens* defendants will be responsible.[366] In essence, a *Bivens* plaintiff faces all the obstacles faced by a § 1983 plaintiff plus more. Mr. Rosen notes that "governmental liability and the right to attorneys' fees, which are not made available to a *Bivens* plaintiff, combined with the extra restrictions applicable only in *Bivens* actions, make the task of the *Bivens* plaintiff that much more difficult than that of an individual suing under Section 1983."[367] Thus, *Bivens* represents an inadequate remedy that fails to provide a plaintiff with any meaningful chance of recovery against a federal official.

V. VIOLATIONS COMMITTED BY SECURITY GUARDS IN PRIVATELY OWNED PLACES OPEN TO THE PUBLIC

As stated earlier, the right to photograph in privately owned places open to the public is governed by relatively straightforward law.[368] Virtually no limits exist on the right to photograph privately owned structures that are visible from public places.[369]

When analyzing the remedies available to plaintiffs harmed by a private security guard, the first inquiry will be whether or not the security guard was acting under color of state law.[370] If the security guard was acting under color of state law, a plaintiff may bring a § 1983 claim.[371] If a court finds that the security guard was not acting under color of state law, a plaintiff must resort to state tort remedies for compensation.[372] Subsection A discusses the requirements for bringing a § 1983 claim against a private

362. *Id.* at 348.
363. *Id.* at 347.
364. *Id.*
365. *Id.*
366. *See supra* pp. 135–37 and accompanying notes.
367. Rosen, *supra* note 333, at 366.
368. *See supra* Section III.B and accompanying notes.
369. *See* R.M.S. Titanic, Inc. v. Haver, 171 F.3d 943, 970 (4th Cir. 1999).
370. *See* 42 U.S.C. § 1983 (2006).
371. *See id.*
372. *See id.*

security guard. Subsection B discusses the remedies available to a plaintiff harmed by a private security guard not acting under color of state law.

A. Private Security Guards Acting under Color of State Law

A private security guard will be considered to be acting under color of state law when his or her conduct is "fairly attributable to the State."[373] In *Chapman v. Higbee Co.*, the Sixth Circuit relied on the Supreme Court's battery of three tests for determining whether a state action exists: "(1) the public function test, (2) the state compulsion test, and (3) the symbiotic relationship or nexus test."[374] The Seventh Circuit has held that private police officers licensed to make arrests could be considered state actors under the public function test.[375] The Sixth Circuit provides further clarification: "Under the public function test, a private entity is said to be performing a public function if it is exercising powers traditionally reserved to the state, such as holding elections, taking private property under the eminent domain power, or operating a company-owned town."[376]

In order for a plaintiff to show that a private security guard acted under color of state law, he or she will generally need to show that the private actor exercised a "power exclusively reserved to the state, e.g., the police power," rather than "a power traditionally reserved to the state, but not exclusively reserved to it, e.g., the common law shopkeeper's privilege."[377] A plaintiff must demonstrate that the private security guard was "endowed by law with plenary police power such that they are *de facto* police officers."[378] If a plaintiff is unable to show that a security guard was a "*de facto* police officer" and his or her actions are not "fairly attributable to the state,"[379] that plaintiff is left with no choice but to resort to common law tort claims for compensation.

Heidi Boghosian writes in the *Missouri Law Review* that private security guards and police officers may be found to be acting under color of law; however, courts have been hesitant to make such a finding.[380] Boghosian states:

> [S]ome courts have ruled that special police status alone does not establish color of law and that the imposition of liability [under § 1983] depends on

373. Lugar v. Edmondson Oil Co., 457 U.S. 922, 937 (1981).

374. Chapman v. Higbee Co., 319 F.3d 825, 833 (6th Cir. 2003) (citations omitted).

375. Payton v. Rush-Presbyterian-St. Luke's Med. Ctr., 184 F.3d 623, 630 (7th. Cir. 1999).

376. Romanski v. Detroit Entm't, L.L.C., 428 F.3d 629, 636 (6th Cir. 2005).

377. *Id.* at 637.

378. *Id.*

379. *See id.* at 636–37.

380. *See* Heidi Boghosian, *Applying Restraints to Private Police*, 70 MO. L. REV. 177, 208–09 (2005).

> whether the officers were performing a "public function"—one traditionally performed for the public good by the state. The Supreme Court has been clear that the scope of public functions is limited, reaching only activities that have been "traditionally the exclusive prerogative of the State."[381]

B. State Tort Claims Available against Private Security Guards not Considered to Be Acting under Color of State Law

As discussed above, if a security guard confiscates a person's camera or unlawfully detains someone, that person will not be able to sue the security guard under § 1983 if the security guard was not "acting under the color of state law."[382] A plaintiff whose rights have been violated must resort to remedies available at state law. Such state tort claims may include, but are not limited to: (1) false imprisonment, (2) false arrest, (3) assault, (4) battery, (5) intentional infliction of emotional distress, (6) interference or conversion of property, and (7) tortuous interference with economic opportunity.[383] The remedies available to plaintiff will be whatever remedies state law provides for each cause of action.[384]

Recall the Chicago incident described in Section II. A girl was standing in line for a concert at the House of Blues and snapped a picture of the security guard, who immediately seized her camera.[385] The girl attempted to take her camera back and the security guard repeatedly shoved her and finally pushed her to the ground.[386] Although the girl may not sue the security guard under § 1983 since he is not acting under the color of state law, she may sue him for use of excessive force, assault, battery, and conversion of her property. Considering the egregious nature of the incident, she would likely succeed on many of these claims.

Ms. Boghosian writes: "Although they perform a range of law enforcement-related activities, private security guards are frequently ill-trained, unsupervised, and may themselves have criminal records."[387] She notes that "the Chicago Housing Authority police chief estimated that 20 percent of guards working private security at the Chicago Housing Authority in 1996 were active gang members."[388] Private security guards "outnumber[] public police by three to one in the United States—in a range of law enforcement activities that put[] them in direct contact with the

---

381. *Id.* at 208 (quoting Rendell-Baker v. Kohn, 457 U.S. 830, 842 (1982)).

382. *See Romanaski*, 428 F.3d, at 636.

383. *See, e.g.*, Bryant, *supra* note 78 (describing incident where plaintiff would have been able to bring a state tort claim against a security guard).

384. *See, e.g., id.*

385. *Id.*

386. *Id.*

387. Boghosian, *supra* note 380, at 177.

388. *Id.*

public."[389] Although in some cases, plaintiffs harmed by private security guards may be adequately compensated by bringing state tort claims, there are a number of obstacles a plaintiff may face.

First, the possibility always exists that the security guard will be insolvent.[390] Second, attorney's fees and court costs may be considerable and render litigation impractical. For example, assume a private security guard confiscates or breaks Person A's camera and deletes the memory card. The court costs of bringing a claim against the security guard may quickly add up to equal or exceed the cost of the camera, and it will be difficult for a plaintiff to receive adequate compensation for the loss of pictures on a memory card because of the difficulty of proving their intended use or estimated worth.

Although plaintiffs suing private security guards pursuant to state torts may face a series of obstacles, these plaintiffs are more likely to be adequately compensated than those bringing a § 1983 suit, for though they will not have access to the deep pockets of the state and will be responsible for court costs and attorney's fees, they will not have to show a constitutional violation nor to circumvent the doctrine of qualified immunity.[391]

VI. SUMMARY OF REMEDIES

A person who turns to the courts for relief when damaged by another's negligent, reckless, or intentional actions can expect to sacrifice time and money to get compensated.[392] This is no surprise. What makes the case of a person damaged within the context of public photography and journalism uniquely unfortunate is that, most of the time, there is no pot of gold at the end of the rainbow.[393] In every step of § 1983 and *Bivens* litigation there are obstacles that often prove insurmountable.[394] Assuming a plaintiff can convince a court of a constitutional violation (which is hindered by courts' ambiguous stance with respect to newsgathering and photographic

389. *Id.*

390. *See id.* at 184 ("The median annual income in 2002 for security guards was $19,140.").

391. *See* Rosen, supra note 333, at 339 (actions under 42 U.S.C. § 1983 involve "constitutional deprivations by government officials" and involve issues of immunity).

392. *See generally id.* Rosen notes additionally that *Bivens* suits have resulted in few plaintiff's verdicts: "Of these [suits], a number have been reversed on appeal and only four judgments have actually been paid by the individual federal defendants." *Id.* at 343–44.

393. *See id.* at 341 (In *Bivens* suits, "[c]ourts and juries [are] therefore left to look to the Park Service policeman, the INS official, or the FBI agent to be financially responsible for the actions each took on behalf of the federal government.").

394. *See id.* at 343 ("The Supreme Court created the *Bivens* doctrine for the express and sole purpose of providing a damages remedy to the victims of constitutional torts. That purpose has simply not been achieved.").

expression), he still must clear the qualified immunity hurdle.[395] Courts apply qualified immunity liberally, even where defendant clearly committed the violation in question.[396] With the help of broadly written criminal statutes, courts are able to find that nearly all official conduct either did not violate the plaintiff's rights or violated them but did so on a reasonable basis.[397]

If an officer behaves so incompetently that he waives qualified immunity, courts may still use causation to limit damages.[398] Damages within the photography and journalism context are difficult to prove,[399] and it is hard if not impossible to put a price on photographs and photographic opportunities. Basically, from beginning to end, § 1983 litigation is packed full of balancing tests, special factors, and subjective notions of what it "reasonable under the circumstances."[400] A court has an unlimited amount of wiggle room and countless opportunities to show deference to police officers.

Bivens actions have all of the bugs (or, if you are a law enforcement officer, features) of § 1983, plus more.[401] Since a *Bivens* action is a judicially created remedy, the court has even greater discretion in limiting the scope of liability and finding "factors that counsel hesitation."[402] Most significantly, courts have not interpreted the *Bivens* remedy to include attorneys' fees.[403]

A plaintiff always has the option of seeking injunctive or declaratory relief, but these remedies come with their own set of problems,[404] the most obvious being their failure to provide pecuniary damages.[405] Securing these forms of relief generally requires extensive factual findings of "patterns of

---

395. *Id.* at 348.

396. *See id.* at 354.

397. *See id.* ("Under the [qualified immunity] test, a federal official who knew he was violating the clearly established constitutional rights of the plaintiff or who acted with malicious intent to violate those rights would still be immune so long as a reasonable official would not have been aware that the actions at issue violated clearly established law.").

398. Ravenell, *supra* note 287, at 722.

399. *See id.* (Some federal courts have applied a strict approach to § 1983 litigation, which "requires that the plaintiff's harm be related to the risk the constitutional amendment was intended to protect.").

400. *Id.* at 692 n.25.

401. *See* Rosen, *supra* note 333, at 357.

402. *Id.*

403. *Id.* at 364.

404. *See* LaDuke v. Nelson, 762 F.2d 1318, 1323 (9th Cir. 1985) ("[W]hen injunctive relief is sought, litigants must adduce a 'credible threat' of recurrent injury."); Evans Med. Ltd. v. Am. Cyanamid Co., 980 F. Supp. 132, 135 (S.D.N.Y. 1997) (finding declaratory relief not available unless a threat is "immediate and real").

405. *See* Rosen, *supra* note 333, at 369 (a "mere violation of constitutional right" does not "automatically result in damages.").

police behavior" and "credible threats" of "recurrent injury."[406] Officers will defend by alleging misconduct or illegal behavior on behalf of the plaintiff (such as "interfering with an investigation" or disturbing the peace).[407] On more than one occasion, courts have taken the officer's word in such a dispute, stating that the First Amendment "does not override the interest of the public," or facilitate "reprehensible conduct forbidden to all other persons."[408]

Practically speaking, the best chance a photographer plaintiff has of securing a meaningful remedy is in a state tort suit against a private security guard. This is because there is no need to establish a constitutional violation and qualified immunity is not at issue. Plaintiffs, however, must bear the financial burden of the litigation, which may be considerable, and also prove damages. A plaintiff's inability to prove the damages he or she suffered could mean a lot of court costs and attorney's fees with little compensation.

In sum, a person has a right to be in a public place, a right to have a camera, and a right to use that camera to photograph persons or structures visible from public places.[409] With few exceptions, federal, state, and local laws do not prohibit such conduct specifically.[410] But when a person's rights in this context are violated, there is no meaningful and practical remedy. The following section discusses why this is a problem and why we should be concerned.

VII. IMPLICATIONS OF AN INADEQUATE REMEDIAL STRUCTURE

In his article *Rights Essentialism and Remedial Equilibration*, Professor Daryl J. Levinson states that, "[r]ights are dependent on remedies not just for their application to the real world, but for their scope, shape and very existence."[411] Indeed, Levinson continues, "it has long been understood that rights and remedies are, in many important contexts, functionally inseparable."[412] The article quotes Oliver Wendell Holmes, who once observed, "[A] legal duty . . . is nothing but a prediction that if a man does or omits certain things he will be made to suffer in this or that way by

406. *See LaDuke*, 762 F.2d at 1323–24.

407. *See* State v. Taylor, 118 A.2d 36, 45 (N.J. Super. Ct. App. Div. 1955).

408. Branzburg v. Hayes, 408 U.S. 665, 691–92 (1971); *see also* Burton, *supra* note 87, at 19 ("Long ago the Court made it clear that the First Amendment does not 'invalidate every incidental burdening of the press that may result from the enforcement of civil and criminal statutes of general applicability.").

409. *See* R.M.S. Titanic, Inc. v. Haver, 171 F.3d 943, 970 (4th Cir. 1999).

410. *See id.*

411. Daryl J. Levinson, *Rights Essentialism and Remedial Equilibration*, 99 COLUM. L. REV. 857, 858 (1999).

412. *Id.*

judgment of the court—and so of a *legal right*."[413] Levinson explains: "'This or that way' is, of course, the remedy."[414] Rights are preserved through remedies because remedies provide a strong incentive for people to recognize and respect those rights. Without remedies, it is unlikely that either federal officials, state officials, or private security guards will be deterred from conduct that violates the rights of others—thus, the right to take photographs in public places will erode.

VIII. WHAT'S WRONG WITH THIS PICTURE?

The erosion of photography rights is a multi-dimensional problem, troubling for a variety of reasons that reflect both substantive and procedural concerns. The first and most obvious concern is that photography is an essential part of the newsgathering process, which is itself essential to a free press. Second and of equal cause for concern is the process by which the right to take pictures in public places has been diminished and the lack of accountability reflected by this process. Lastly, the erosion of photography rights has not been accompanied with a corresponding increase in national security or public safety, nor will it be in the future. This section discusses each of these three problems in some detail.

A. The Erosion of Photography Rights Will Diminish Public Awareness, Transparency, and Accountability

Almost everyone is familiar with the adage "A picture is worth a thousand words." A 1998 *Harvard Law Review* note cites two particular instances in American history demonstrating the power of photographs.[415] First, it mentions the picture depicting the aftermath of the infamous 1970 shooting at Kent State University.[416] On May 4, 1970, the Ohio National Guard opened fire on a large group of student protesters without warning.[417] The photograph, taken by John Paul Filo, shows a young woman kneeling over the body of a dead student.[418] The look on her face is one of disbelief and horror, reflecting the prevailing sentiment of those who experienced the massacre.[419] Within days, the image appeared on the front pages of

413. *Id.* (citing Oliver Wendell Holmes, *The Path of the Law*, 10 HARV. L. REV. 457, 458 (1897)) (emphasis added).

414. *Id.*

415. *See* Note, *supra* note 102, at 1095 (citing VICKI GOLDBERG, THE POWER OF PHOTOGRAPHY: HOW PHOTOGRAPHS CHANGED OUR LIVES 7 (1991)).

416. *See id.* at 1095.

417. *See id.*

418. *See id.*

419. *See id.*

newspapers across the country.[420] Vicki Goldberg, a photography critic and author, notes that "'[t]he photograph helped galvanize the stalled antiwar movement on college campuses and came to symbolize a nation's shock that its children were dying at the hands of its protectors.'"[421] The second photograph, taken by Huynh Cong Ut, captured the horror of the Vietnam War,[422] showing several children in South Vietnam frantically fleeing a napalm strike.[423] The central figure is a young girl "stark naked and badly burned; she had torn off all of her clothes in a futile attempt to escape the searing effects of napalm."[424] Tremendously disturbing, this photograph became an icon of the antiwar movement.[425] Susan Sontag, author of a collection of essays on the history and present-day role of photography, stated that Ut's photograph "'probably did more to increase the public revulsion against the war than a hundred hours of televised barbarities.'"[426]

These images are striking examples of the power of photography to increase public awareness, shape the nation's conscience, inspire change, and instantaneously convey volumes of information in a manner often more effective than through the written or spoken word. As Goldberg notes, "'[p]hotographs have a swifter and more succinct impact than words, an impact that is instantaneous, visceral, and intense.'"[427] A powerful image is simply harder to ignore than a written or spoken description of the same image. One reason for this is that an image's effectiveness in conveying an idea does not depend as much on the viewer's trust in the source. The viewer has the ability to see for him or herself, making it harder to dismiss the information on the basis of questionable validity. For these and other reasons, photography represents a powerful tool for increasing public awareness and inspiring reform.

Photography is also a valuable means of enhancing accountability on behalf of law enforcement officials and private security guards. The mere awareness that one's behavior can be captured on camera provides a powerful incentive for officials to avoid acting outside the scope of their authority. Additionally, when police misconduct does occur and is captured on film, publication of the footage inspires public outrage and, consequently, reform (or at least increased vigilance). Take for example the reaction to the police beating of Rodney King, captured on camera in March 1991.[428] An individual captured Los Angeles Police Department

420. *See id.*

421. *Id.* (quoting GOLDBERG, *supra* note 415, at 237).

422. *See id.*

423. *See id.*

424. *Id.* at 1095–96.

425. *Id.*

426. *Id.* (citing SUSAN SONTAG, *In Plato's Cave*, in ON PHOTOGRAPHY 3, 18 (1977)).

427. Note, *supra* note 102, at 1095 (quoting GOLDBERG, *supra* note 415, at 7).

428. *See* Elizabeth F. Loftus & Laura A. Rosenwald, *The Rodney King Videotape: Why the Case Was Not Black and White*, 66 S. CAL. L. REV. 1637, 1637 (1993).

officers repeatedly striking King with their batons.[429] A portion of the footage was aired by news agencies around the country, causing public outrage over police brutality.[430]

The foregoing illustrates that the right to take pictures in public places is an important one not only because of its communicative and expressive value, but also because it helps ensure the survival of other rights by promoting accountability and transparency. Unfortunately, much of the usefulness of photography in this capacity depends on the ability of photographers to take pictures without the burden of undue governmental regulation. When police officers unlawfully prohibit a photographer from taking a picture, the individual photographer is harmed, but more importantly, the incident contributes to a chilling effect on photography in general. A photographer who fears that taking a picture may result in criminal charges and that few if any legal remedies are available may decide that taking the picture is not worth the risk. The value of photography to individuals and to society cannot be fully realized if those holding the camera are forced to think before they snap.

B. The Erosion of Photography Rights Reflects Insulated Policymaking Void of Accountability

The process by which photography rights have been diminished is equally troubling as the concerns discussed in the previous section. The so-called "War on Photography" is not the product of a calculated effort on behalf of policymakers designed to enhance national security or public safety. Instead, it seems to be the result of haphazard snap judgments on part of law enforcement officials and security guards, which seem to be motivated either by the belief that they are promoting national security or public safety, or in some cases, a desire to avoid having their behavior caught on camera. Their source of authority? Generally, they point either to a fictitious "national security" or "9/11" law or a broadly worded statute the drafters of which probably never intended to apply to the photography in question. Because of the lack of adequate remedies and the courts' reluctance to entertain claims under the remedies that do exist, law enforcement officials have no incentive to avoid interfering with lawful behavior.

Undoubtedly, in many cases, police officers are simply trying to promote public safety and do their job. The problem is that when an officer determines that a person should not be able to photograph a skyscraper, a bridge, or an ongoing arrest, the officer is not doing *his* job—he is doing the job of the legislature. It is the responsibility of the legislature to determine which behaviors are unlawful; the province of law enforcement officers is determining whether or not a particular person's actions constitute the

429. *See id.*
430. *See id.*

behavior proscribed by the legislature. As stated throughout this article, few laws exist prohibiting photography in public places, and the ones that do prohibit only the most obnoxious or intrusive behavior. Thus, when an officer unilaterally determines that taking pictures in a particular instance is prohibited, he abandons the role of a law enforcer and assumes the role of policymaker. When unelected officials such as law enforcement officials make policy judgments they are not entitled to make, the link between citizen and policymaker is severed and accountability is destroyed. To make matters worse, the inadequate remedial structure within the realm of photography rights combined with courts' deferential treatment to law enforcement officials effectively insulate officers' conduct from consequences. Law enforcement officials have little incentive to refrain from such conduct and photographers have few avenues available to them to initiate change.

C. The Erosion of Photography Rights Has not Been and Will not Be Accompanied by an Increase in National Security or Public Safety

Undue regulation of the right to take photographs in public is not only intrusive, burdensome and limiting; it is also ineffective at enhancing national security or public safety. Modern technology has made it possible to access online images of buildings, transportation stations and a multitude of other structures from a variety of angles and distances.[431] What is the advantage of limiting a person's ability to photograph structures when the same or similar images are widely available to the general public? Additionally, evidence exists that the perpetrators of past terrorist plots never photographed their targets. Bruce Schneier, an internationally renowned security technologist and author noted:

> The 9/11 terrorists didn't photograph anything. Nor did the London transport bombers, the Madrid subway bombers, or the liquid bombers arrested in 2006. Timothy McVeigh didn't photograph the Oklahoma City Federal Building. The Unabomber didn't photograph anything; neither did shoe-bomber Richard Reid. Photographs aren't being found amongst the papers of Palestinian suicide bombers. . . . Even those manufactured terrorist plots that the US government likes to talk about—the Ft. Dix terrorist, the JFK airport bombers, the Miami 7, the Lackawanna 6—no photography.[432]

Schneier argues that even if terrorists did photograph their targets, it would not be practical to prevent it:

---

431. *See, e.g.*, GOOGLE EARTH, http://www.google.com/earth/index.html (last visited Jan. 25, 2011).

432. Schneier, *supra* note 2.

> Billions of photographs are taken by honest people every year, 50 billion by amateurs alone in the US. And the national monuments you imagine terrorists taking photographs of are the same ones tourists like to take pictures of. If you see someone taking one of those photographs, the odds are infinitesimal that he's a terrorist.[433]

Indeed, one can argue that photography enhances rather than hinders national security. In a *Popular Mechanics* article entitled *Photo Phobia*, University of Tennessee College of Law Professor Glenn Reynolds writes: "Even in potential terrorism cases, the presence of lots of ordinary folks carrying cameras actually enhances public security. In the hours after the failed Times Square car-bomb attempt, officials searching for clues didn't just look at their own security-camera footage, they also sought out home movies shot by tourists."[434] Ironically, then, law enforcement officials' efforts to enhance security by limiting photography are likely to have exactly the opposite effect.

The following section of this article proposes solutions for preserving the right to take pictures in public places and ideas to prevent further violations of these critical rights. Some of the solutions discussed are prophylactic while others are more remedial in nature.

IX. POSSIBLE SOLUTIONS

A. Know Your Rights

The first way photographers can prevent abuse is to know their rights. When confronted by law enforcement officers, photographers who do not know and understand their rights are more likely to apologize for their conduct and comply with authority, whether or not this authority is legitimate. As a consequence, Bruce Schneier writes, "[l]aw enforcement officials and security guards are then emboldened to enforce a nonexistent law and trample on constitutional rights, and there is no incentive for them to do otherwise."[435] Schneier advises photographers, journalists, and others to carry cards or pamphlets listing their legal rights and obligations in the event they are confronted by law enforcement officials for taking photographs.[436] Attorney Bert P. Krages also advocates the use of pocket guides and provides one on his website.[437] He advises photographers to

---

433. *Id.*

434. Glenn Harlan Reynolds, *Photo Phobia*, POPULAR MECHANICS, Aug. 2010, at 52–53.

435. Bruce Schneier, *The War on Photography*, SCHNEIER ON SECURITY (June 5, 2008, 6:44 AM), http://www.schneier.com/blog/archives/2008/06/the_war_on_phot.html.

436. Bruce Schneier, *Are Photographers Really a Threat?*, THE GUARDIAN, June 5, 2008, http://www.guardian.co.uk/technology/2008/jun/05/news.terrorism.

437. Bert P. Krages, *The Photographer's Right*, *available at* http://www.krages.com/ThePhotographersRight.pdf.

carry the guide in their wallet or camera bag for "quick access to your rights and obligations concerning confrontations over photography."[438]

B. Publicize Violations

Organized media at every level (local, state and national) should shed light on incidents where photography rights are curbed. Throughout this article there are a number of examples of how media organizations have the power to publish stories that serve as a catalyst for change. Consistent publication of violations of photographers' rights will increase public awareness of the issue and motivate people to act. Increased attention may also motivate law enforcement brass to be more vigilant in monitoring their officers' behavior and discipline them for acting outside their authority.

News organizations should ensure that journalists understand their right to photograph and inform journalists of the relatively few restrictions that exist on those rights. They should advise journalists and photographers that if they feel they are about to be the victim of unlawful police conduct and believe themselves to be acting within the scope of the law, they should continue taking photographs or filming what is happening. The photos or footage may prove to be valuable evidence in litigation and an effective tool for educating the public and inspiring action. Lastly, media organizations should provide attorneys for photographers charged by police and indemnify them on any judgments rendered.[439]

C. Attorneys: Attack the Statute

It is important for attorneys who defend photographers and journalists to include in their defense an attack on the statute or regulation itself. Counsel should not limit their position to a purely factual defense. In addition to this argument, counsel should attack the statute itself on the basis that it criminalizes constitutionally protected behavior and is being unconstitutionally applied to press. Counsel should attack broadly written statutes as unconstitutionally vague or overbroad and argue that these statutes give law enforcement officials unlimited discretion to target press.

Attacking the statute itself represents an efficient way to combat police misconduct, because it works to eliminate one of law enforcement officials' primary tools for suppressing journalists and photographers. It also increases courts' awareness that such "all-purpose" statutes are being used to target the organized media and hinder the newsgathering process. Attacking the statute also makes it more likely that you will achieve your client's goals and also helps prevent future abuse of photography rights.

---

438. Bert P. Krages, *The Photographer's Right*, BERT P. KRAGES, ATTORNEY AT LAW PHOTOGRAPHER'S RIGHTS PAGE, http://www.krages.com/phoright.htm (last visited Jan. 25, 2011).

439. *See* Burton, *supra* note 87, at 22.

D. An Organized Front on Behalf of News Media

The third solution is for media organizations at all levels (local, state and national) to dedicate resources to developing an organized front to combat abuses of newsgathering and press rights. Eve Burton, Vice President and General Counsel of the Hearst Corporation, argues that the organized media's reaction to violations of newsgathering and press rights has been weak.[440] Her article *Where Are All the Angry Journalists?* argues that the organized media and the general public have not demonstrated an appropriate level of outrage at the restriction of press rights.[441]

In addition to investing in able attorneys to represent journalists in civil actions, Burton proposes that the organized media create a national or state-by-state data bank of information regarding violations of newsgathering and press rights.[442] With this information at their fingertips, counsel for media organizations can make a more convincing and well-documented argument that the statute in question has been used to target the constitutionally protected activities of the organized media. Additionally, a data bank could be useful in equipping media lobbyists with strong evidence that could be used to convince legislators to repeal or amend existing statutes.

E. Observe the Efforts of Photographers Abroad

The efforts of journalists and photographers in other countries provide helpful guidance in applying pressure to federal officials in the United States. In London, *Amateur Photographer* magazine, the Royal Photographic Society and the British Institute of Professional Photography have all been instrumental in influencing the government to take measures to prevent the abuse of press rights.[443] In March 2009, the Home Office invited representatives of the media to "help draft guidance that will aim to ensure police do not misuse anti-terrorism legislation to unfairly stop photographers."[444] This invitation came in response to a request by *Amateur Photographer* that Parliament "adopt a common-sense approach when dealing with photographers."[445] In addition to efforts by the organized media, hundreds of photographers staged a demonstration outside New Scotland Yard to protest broad anti-terrorism legislation that they felt would be used to target the lawful activities of photographers.[446]

---

440. *Id.* at 19.

441. *Id.* at 21–22.

442. *Id.*

443. Chris Cheesman, *Photographers Meet home Office Minister Over Rights "Abuse,"* AMATEUR PHOTOGRAPHER, Mar. 12, 2009, http://www.amateurphotographer.co.uk/news/photographers_meet_home_office_minister_over_rights_abuse_news_278619.html.

444. *Id.*

445. *Id.*

446. *Id.*

F. Expand the § 1983 Definition of "Acting under Color of State Law" to Include Private Security Guards

In response to abuse by private security guards, media organizations and other groups should pressure Congress and the courts for an expanded § 1983 definition of "acting under state law."[447] Security guards stop, detain and search people just like police. They wear uniforms similar to those of police and carry the same weapons police carry. As Heidi Boghosian's observes, "[T]he private police industry relies on uniforms to imbue its agents with the public police's implied monopoly on state-sanctioned use of force and coercion . . . . As a result, private security personnel are frequently mistaken for public police."[448]

As discussed in Section V., courts are hesitant to make a finding that a privacy security guard "acted under color of state law." Attorneys for media organizations need to argue for an interpretation of "acting under color of state law" that more accurately reflects the reality that police officers and private security guards often assert the same authority and are in many cases indistinguishable. In addition, attorneys should encourage courts to fashion equitable remedies under § 1983 in the form of improved training and oversight of security guards.[449] Improved training and increased accountability will help decrease future abuses by security guards.

G. New Legislation under the Fourteenth Amendment

The Fourteenth Amendment to the United States Constitution empowers Congress to pass laws that protect the civil rights of citizens.[450] Media organizations and other groups affected by police misconduct in the realm of photography rights should pressure Congress to adopt legislation aimed at preventing the abuses of the variety described in this article. Professor Glenn Reynolds considers the abuse of rights in the realm of photography to be "one of the comparatively few issues that could merit a new federal civil rights law."[451] Professor Reynolds believes that a clear federal law would limit situations in which "local officials use their power to harass those who might keep an eye on them. Passing such a law would make us all safer."[452]

---

447. *See* Boghosian, *supra* note 380, at 208–09 (describing uncertainty regarding whether courts will treat security guards as "acting under color of state law").

448. *Id.* at 204.

449. *Id.* at 211.

450. U.S. CONST. amend. XIV, § 5.

451. Reynolds, *supra* note 434.

452. *Id.*

X. CONCLUSION

The law in the United States places few restrictions on the right to take pictures in public places.[453] Yet since 9/11, photographers have repeatedly been harassed, questioned, detained, and charged with crimes. The War on Terror has augmented law enforcement's already distrusting and hostile attitude towards photographers. Police officers have repeatedly charged or threatened to charge journalists pursuant to broadly worded criminal statutes and vague references to national security.[454] Photographers have had their cameras taken unlawfully, memory cards deleted, and their dignity sullied by police officers and private security guards acting outside of their authority. Worse yet, photographers whose rights have been violated have few, if any, meaningful remedies available to them under existing law. All of the remedies discussed in this paper are difficult to obtain and unlikely to compensate a plaintiff fully. In most cases, the chance of recovery is slim, and the amount of recoverable damages is uncertain. Considering the time and resources associated with litigation, it is often impractical for a victim to pursue legal remedies for violations of his or her photography rights.

Without meaningful remedies, the right to take pictures in public places will further erode. Photography is a powerful mode of expression and a valuable communicative tool. It is an essential part of a free press, which is itself essential to democracy. Violations of rights within the context of newsgathering and photography indubitably cause harm to the individual. But more importantly, these violations cause significant damage to society by threatening the free flow of information and ideas.

Ameliorating what scholars have labeled "the War on Photography" will require the efforts of both the organized media and grassroots organizations. Both should seek to educate people about their legal rights and increase the public's awareness of abuses described. The organized media needs to dedicate the resources necessary to defend those charged with crimes and to pressure courts and legislatures to make and interpret law in a way that holds officials accountable for their actions.

These and other efforts are needed to prevent the continuation of widespread abuse of photography and newsgathering rights. The meaningful remedies needed to preserve the right to free press and expression are currently lacking. In the words of one scholar, "[t]he death of a free press can occur not only by a dictator's edict but by slow erosion, one case at a time."[455] If efforts are not made to resolve the current problem,

---

453. Note, *supra* note 102, at 1088.

454. *See generally* Burton, *supra* note 87 (providing examples of journalists being charged under broadly worded criminal statutes).

455. Burton, *supra* note 87, at 22 (citing John Grogan, *A Case of Ethics vs. Pursuit of Justice*, SUN-SENTINEL, Oct. 11, 1996, at 1B.).

both individuals and society at large will suffer. Freedom of press and expression will indeed be diminished "one case at a time."[456]

456. *Id.*

COPYRIGHTING COUTURE: AN EXAMINATION OF FASHION DESIGN PROTECTION AND WHY THE DPPA AND IDPPPA ARE A STEP TOWARDS THE SOLUTION TO COUNTERFEIT CHIC

SARA R. ELLIS*

"It is clear Congress intended the scope of the copyright statute to include more than the traditional fine arts."[1]

I. INTRODUCTION 164
II. BACKGROUND 166
A. *The Origin of Design Piracy in the Fashion Industry* 169
B. *The Current Status of United States Law and Fashion Design* 169
1. Intellectual Property under the Constitution 170
2. Copyright Law 170
3. Trademark & Trade Dress Law 174
a. *Trademarks* 174
b. *Trade Dress* 176
4. Patent Law 178
C. *Prior Congressional Consideration & Attempts to Obtain Protection for Fashion Design* 179
1. The Design Registration League 180
2. Fashion Originators' Guild of America 180
3. H.R. 5055: To Amend Title 17, United States Code, to Provide Protection for Fashion Design, 109th Congress ... 182
4. H.R. 2033 & S. 1957: Design Piracy Prohibition Act, 110th Congress 183
5. H.R. 2196 Design Piracy Prohibition Act & S. 3728 Innovative Design Protection and Piracy Prevention Act, 111th Congress 184
III. ANALYSIS 185
A. *Why Is Fashion Design Protection Necessary?* 185

---

* Candidate for Doctor of Jurisprudence, May 2011, University of Tennessee, College of Law; Research Editor, *Tennessee Law Review*; Third Year Editor, *Transactions: The Tennessee Journal of Business Law*; Saul Lefkowitz Trademark Moot Court Team. The author would like to thank Professor Gary Pulsinelli and Dean Carol Parker for their insight, guidance, and contributions. The author would also like to give special thanks to Millard Powers for his continuous support.

1. Mazer v. Stein, 347 U.S. 201, 213 (1954).

1. Fashion is Money 185
2. Fashion is a Form of Art and Should Receive Protection Equivalent to the Traditional Fine Arts 186
3. Piracy Affects Designers, Large and Small 188
4. Similar Legislation in the European Union has Aided Fashion Designers in Protecting their Designs 191
5. Similar Legislation in the European Union Does Not Indicate the Likelihood of Decreased Creativity or an Impeded Design Cycle 194
6. Similar Legislation in the European Union Does Not Indicate the Likelihood of Increased Litigation 194

B. What Do The DPPA and IDPPPA Mean? 195
1. Limitations on Designs Protected 196
a. Similarities Between the DPPA and IDPPPA 196
b. Additions Proposed by the IDPPPA 197
c. Proposed Compromise 198
2. Revisions, Adaptations, and Rearrangements 199
3. Publication of Designs 200
4. Term of Protection 202
5. Registration 202
a. The DPPA 202
b. The IDPPPA 203
c. Proposed Compromise 203
6. Infringement 204
a. Acts with Knowledge 204
b. Infringing Article Defined 205
c. Secondary liability 208
d. Home Sewing Exception—IDPPPA Only 208
7. Recovery for Infringement—*DPPA* Only 208
8. Remedy for Infringement—*IDPPPA* Only 209
9. Effective Date 210

IV. CONCLUSION 210

I. INTRODUCTION

In recent years, there has been much debate about the need for intellectual property protection for fashion designs. Two bills introduced before the 111th Congress purported to provide a solution for this need.[2] The Design Piracy Prohibition Act ("DPPA"), which was introduced in the House April 30, 2009,[3] and the Innovative Design Protection and Piracy

2. *See* S. 3728, 111th Cong. (2010); H.R. 2196, 111th Cong. (2009).
3. H.R. 2196, 111th Cong. (2009).

Prevention Act ("IDPPPA"), which was introduced in the Senate on August 5, 2010,[4] would have amended Chapter 13 of the Copyright Act[5] to provide *sui generis*[6] protection for fashion designs.[7] While some scholars worry that bills like the DPPA and the IDPPPA would stifle creativity,[8] many designers think that such protection would afford them more freedom to create new and innovative designs.[9] Scholarly debate aside, piracy is a $12 billion[10] drain on the fashion industry that "steals the very essence of enterprise"[11] by diluting branding and making it more difficult for new designers to begin their careers.[12]

Because current intellectual property laws do not address the unique issues involved in fashion design, pirates appropriate, or even directly replicate, others' designs even in the face of a constant stream of lawsuits.[13] For example, the company Forever 21, one of the most notorious design thieves, was the subject of over fifty lawsuits between 2006 and 2009 alone.[14] The retail chain is a billion-dollar-per-year enterprise that sells discount clothing to young women and thrives by quickly copying others' designs.[15] Because of Forever 21's practice of copying, the company has been sued by Trovata,[16] Diane Von Furstenberg,[17] Anna Sui,[18]

---

4. S. 3728, 111th Cong. (2010).

5. *See* Copyright Act, 17 U.S.C. §§ 101–1332 (2006).

6. *Sui generis* is Latin for "of its own kind" and applies to intellectual property areas "that fall outside the traditional patent, trademark, copyright, and trade-secret doctrines." BLACK'S LAW DICTIONARY 1572 (9th ed. 2009).

7. *See, e.g.*, S. 3728, 111th Cong. (2010); H.R. 2196, 111th Cong. (2009).

8. *See A Bill to Provide Protection for Fashion Design: Hearing on H.R. 5055 Before the H. Subcomm. on Courts, the Internet, and Intellectual Prop. of the H. Comm. on the Judiciary*, 109th Cong. 87–88 (2006) [hereinafter *H.R. 5055 Hearing*] (statement of Christopher Sprigman, professor at the University of Virginia School of Law).

9. *See* STOP FASHION PIRACY, http://www.stopfashionpiracy.com (last visited Jan. 1, 2011) (video on main page of website).

10. See *Design Law—Are Special Provisions Needed To Protect Unique Industries?: Hearing on H.R. 2033 Before the H. Subcomm. on Courts, the Internet, and Intellectual Prop. of the H. Comm. on the Judiciary*, 110th Cong. 22 (2008) [hereinafter *H.R. 2033 Hearing*] (statement of Narciso Rodriguez, an American fashion designer).

11. STOP FASHION PIRACY, *supra* note 9.

12. *See id.*

13. *See* Amy Odell, *Forever 21's Ability to Copy Designer Clothes Could Be in Jeopardy*, N.Y. MAG. (Apr. 13, 2009, 9:45 AM), http://nymag.com/daily/fashion/2009/04/forever_21s_ability_to_copy_de.html.

14. *See id.*

15. *See* Ruth La Ferla, *Faster Fashion, Cheaper Chic*, N.Y. TIMES, May 10, 2007, at G2, *available at* http://www.nytimes.com/2007/05/10/fashion/10FOREVER.html.

16. *See* Odell, *supra* note 13.

17. *See* First Amended Complaint, Diane von Furstenberg Studio, L.P., v. Forever 21, Inc., No. 07 CV 2413 (VM) (S.D.N.Y. Apr. 12, 2007), 2007 WL 1643831.

18. *See* Complaint, Anna Sui Corp. v. Forever 21, Inc., No. 07 Civ. 7873(RJS)(MHD), No. 07 CV 3235(TPG) (S.D.N.Y. Apr. 23, 2007), 2008 WL 4386747.

Anthropologie,[19] Gwen Stefani,[20] Express,[21] and many others. Their claims have included trademark infringement, trademark dilution, willful copyright infringement, unfair competition and false designation of origin or sponsorship, and unlawful deceptive acts and practices.[22] Unfortunately, intellectual property law's current status makes it very difficult for designers to find relief for their pirated designs.[23] Forever 21 settles most of its cases,[24] so designers rarely get the satisfaction of holding the company accountable in a court of law.[25] Forever 21 is just one example of the companies that use this business practice,[26] but it clearly represents the flaws in the current law.

This Comment discusses the piracy problems plaguing the fashion industry and offers a potential solution based on combining the best aspects of the DPPA and the IDPPPA. Part II examines design piracy—both what it is and the attempts that have been made to gain protection against it—and discusses why current jurisprudence insufficiently addresses the needs of the American fashion industry. Part III.A analyzes the need to protect fashion designs and addresses opponents' concerns. In light of those concerns, Part III.B examines how each bill addresses design piracy and suggests potential changes that would make each bill more effective at protecting fashion designs. Finally, Part IV offers parting commentary.

II. BACKGROUND

Piracy is "[t]he unauthorized and illegal reproduction or distribution of materials protected by copyright, patent, or trademark law."[27] In the fashion industry, piracy takes the form of design copying, which generally falls into one of three categories: trend imitation,[28] counterfeit goods,[29] and style piracy or knockoffs.[30]

---

19. *See* Complaint and Jury Demand, Anthropologie, Inc., v. Forever 21, Inc., No. 07 CV 7873(RJS)(MHD) (S.D.N.Y Sept. 6, 2007), 2009 WL 1383605.

20. *See* Complaint for False Designation of Origin, Trademark Infringement, Dilution and Unfair Competition, Harajuku Lovers, L.L.C, v. Forever 21, Inc., No. CV-07-3881-ODW (SSx) (C.D. Cal. June 14, 2007), 2008 WL 5973843.

21. *See* Laurel Pinson, *Express vs. Forever 21*, NBC WASH. (June 29, 2009, 7:09 PM), http://www.nbcwashington.com/around-town/fashion/Express-vs-Forever-21.html.

22. *See* Complaint, *supra* note 18, at 12–13; Complaint for False Designation of Origin, Trademark Infringement, Dilution and Unfair Competition, *supra* note 20, at 5–9; First Amended Complaint, *supra* note 17, at 7–9.

23. *See* Rachel Brown, *Trovata, Forever 21 Return to Square One*, WOMEN'S WEAR DAILY (May 29, 2009, 11:33 AM), http://www.wwd.com/fashion-blogs/trovata_forever_21_return_to_s-09-05.

24. *See* Odell, *supra* note 13.

25. *See* Brown, *supra* note 23.

26. *See* La Ferla, *supra* note 15, at G5.

27. BLACK'S LAW DICTIONARY 1266 (9th ed. 2009).

28. *See* C. Scott Hemphill & Jeannie Suk, *The Law, Culture, and Economics of*

Trend imitation is the "trickle-down effect" from haute couture[31] to the local Casual Corner.[32] Here, a designer references the works of others, but does not directly copy them.[33] The original pieces are used as a source of inspiration for the designer's own sense of expression.[34] This includes the "hot" color of the season, the hem length of skirts, and the general cut of designs.[35] This form of imitation is not considered infringement; it illustrates a distinction between those who simply replicate the work of others and those who contribute their own creativity.[36]

In contrast, counterfeit goods are illegal goods that are passed off as original products but do not come from the original source.[37] Because they

---

Fashion, 61 STAN. L. REV. 1147, 1160 (2009).

29. *See* Silvia Beltrametti, *Evaluation of the Design Piracy Prohibition Act: Is the Cure Worse than the Disease? An Analogy with Counterfeiting and a Comparison with the Protection Available in the European Community*, 8 NW. J. TECH. & INTELL. PROP. 147, 147 (2010).

30. *See* Kristin L. Black, Comment, *Crimes of Fashion: Is Imitation Truly the Sincerest Form of Flattery?*, 19 KAN. J.L. & PUB. POL'Y 505, 510 (2010) (citing Fashion Originators Guild of Am. v. Fed. Trade Comm'n, 312 U.S. 457, 461 (1941)).

31. Haute couture is French for "high sewing" and generally refers to trend-setting fashions. *See* THE OXFORD ENGLISH DICTIONARY 1067 (2d ed. 1991).

32. A scene in the movie *The Devil Wears Prada* illustrates how trends make their way to even the most humble consumer. In the scene, Miranda Priestly, a character based on Vogue editor in chief Anna Wintour, explains to her assistant how the decisions made in her office affect what ends up in the assistant's closet:

> You go to your closet and you select out . . . I don't know. . . that lumpy blue sweater, for instance, because you're trying to tell the world that you take yourself too seriously to care about what you put on your back. But what you don't know is that that sweater is not just blue, it's not turquoise, it's not lapis, it's actually cerulean. You're also blithely unaware of the fact that in 2002, Oscar de la Renta did a collection of cerulean gowns. And then, I think it was Yves Saint Laurent, wasn't it? Who showed cerulean military jackets? . . . And then cerulean quickly showed up in the collections of eight different designers. And then it filtered down through the department stores. And then it . . . trickled on down into some tragic Casual Corner where you, no doubt, fished it out of some clearance bin. However, that blue represents millions of dollars and countless jobs, and it's sort of comical how you think that you've made a choice that exempts you from the fashion industry when, in fact, you're wearing a sweater that was selected for you by the people in this room from a pile of stuff.

THE DEVIL WEARS PRADA (Fox 2000 Pictures 2006).

33. *See* Hemphill & Suk, *supra* note 28, at 1160.

34. *See id.*

35. *See id.* at 1159–60.

36. *See id.* at 1159–61.

37. *See* BLACK'S LAW DICTIONARY 402–03 (9th ed. 2009) ("Counterfeiting includes producing or selling an item that displays a reproduction of a genuine trademark, usu[ally] to

are illegal, these items are often sold in the back rooms of businesses, although some brazen vendors sell them in conspicuous places like Canal Street in New York or Santee Alley in Los Angeles.[38]

Similarly, knockoffs are intended to replicate the original nearly line for line,[39] but with another designer's name attached.[40] Some consider it "design poaching"[41] or "counterfeiting without the label."[42] This problem can affect anyone, from a start-up designer who finds replicas of his or her hand-made designs on sale at Abercrombie & Fitch,[43] to a famous designer of haute couture gowns whose designs are replicated by notorious knockoff artist Allen Schwartz just hours after they appear on the red carpet.[44] In the instance of designer purses, a knockoff occurs when the designer's logo is altered to be more generic while maintaining the product's same overall look.[45] For example, a Chanel bag may be replicated line for line using double-Os as the logo instead of the interlocking, back-to-back Cs.[46] These knockoffs are typically targeted at "people who appreciate high style but can't afford high prices."[47] Products in this third category are currently legal and will be the focus of this Comment.[48]

---

deceive buyers into thinking they are purchasing genuine merchandise.").

38. *See* Andrea Sachs, *School of Hard Knock-Offs*, WASH. POST (Sept. 4, 2005), http://www.washingtonpost.com/wpdyn/content/article/2005/09/02/AR2005090200826.html (discussing counterfeit goods sold on New York's Canal Street); Alex Schmidt, *Canal Knockoffs Move to the Backrooms*, DOWNTOWN EXPRESS, Feb. 3, 2006, *available at* http://www.downtownexpress.com/de_143/canalknockoffsmovetothe.html (discussing counterfeit goods being moved to the back rooms of businesses); Emili Vesilind, *The New Pirates*, L.A. TIMES, Nov. 11, 2007, *available at* http://articles.latimes.com/2007/nov/11/image/ig-piracy11 (discussing counterfeit goods sold on the streets of Los Angeles).

39. A line-for-line copy is a nearly identical imitation of an original design. *See H.R. 5055 Hearing*, *supra* note 8, at 180 (statement of Susan Scafidi, professor at Fordham Law School).

40. *See* Vesilind, *supra* note 38 ("'There is no counterfeiting without design piracy,' designer Diane von Furstenberg said in an interview at her Beverly Hills estate. 'It's counterfeiting without the label.'").

41. Amy Kover, *That Looks Familiar. Didn't I Design It?*, N.Y. TIMES, June 19, 2005, at B4, *available at* http://www.nytimes.com/2005/06/19/business/yourmoney/19bags.html.

42. *See* Vesilind, *supra* note 38.

43. *See* Kover, *supra* note 41.

44. *See Oscar Dresses from the Red Carpet and Cheap Chic Celebrity Inspired Looks*, FOCUSONSTYLE.COM (Mar. 12, 2010), http://www.focusonstyle.com/Trends/Fashion/Oscar-dresses-red-carpet-copy-cat-cheap-chic-2010.

45. *See* Vesilind, *supra* note 38.

46. Vesilind, *supra* note 38 ("'If there's a copy of a Chanel bag that has a logo that's double-Os instead of double-Cs, we can't do anything, . . . [b]ut once the vendor cuts the Os into a double-C logo, that's a violation.'").

47. James Surowiecki, *The Piracy Paradox*, NEW YORKER, Sept. 24, 2007, at 90, *available at* http://www.newyorker.com/talk/financial/2007/09/24/070924ta_talk_surowiecki.

48. *See H.R. 5055 Hearing*, *supra* note 8, at 4 (statement of Rep. Bob Goodlatte).

A. The Origin of Design Piracy in the Fashion Industry

The cry for intellectual property protection for fashion designs is not new.[49] Silk weavers in Lyon, France first demanded protection for their designs in the early 1700s, and "by 1787 a royal decree extended the protection to silk manufacturers nationwide."[50] In the 1850s, French designer Charles Worth initiated the practice of releasing new clothing designs each season, just as designers do today.[51] Unfortunately, this practice "spawned an industry of knockoff artists" who reproduced the Parisian originals for a much lower price.[52] The French, always on the cutting edge of fashion design and design protection, responded to this piracy in two ways: "[F]irst, by seeking intellectual property protection for original fashion designs[,]" and second, by licensing their original designs to other manufacturers, who then reproduced them line for line.[53] Those able to afford the high cost of the haute couture flocked to Paris to be the first to own the newest designs, while the middle class purchased legally licensed copies from their local department stores.[54]

Professor Susan Scafidi[55] opines that although the French intellectual property laws did not completely eradicate piracy, their protection led to the strength of the fashion industry in Paris and has helped it continue to thrive today.[56] While France's fashion industry flourished, "the United States instead became a haven for design pirates who strenuously resisted efforts to introduce laws protecting fashion."[57]

B. The Current Status of U.S. Law and Fashion Design

Current intellectual property laws are woefully inadequate to meet the needs of the American fashion industry.[58] This section will first discuss the genesis of intellectual property law in the context of the United States Constitution. Next, it will discuss the different areas of applicable

---

49. *See, e.g.*, Susan Scafidi, *Intellectual Property and Fashion Design*, *in* 1 INTELLECTUAL PROPERTY AND INFORMATION WEALTH 115, 116 (Peter K. Yu ed., 2006).

50. *Id.* (citing JEANNE BELHUMEUR, DROIT INTERNATIONAL DE LA MODE 70 (2000)).

51. *See id.* (citing PHILIPPE PERROT, FASHIONING THE BOURGEOISIE: A HISTORY OF CLOTHING IN THE NINETEENTH CENTURY 184–88 (Richard Bienvenue trans., 1994)).

52. *Id.*

53. *Id.*

54. *See id.*

55. Susan Scafidi teaches fashion law at Fordham Law School and testified in the congressional hearing regarding H.R. 5055, A Bill to Provide Protection for Fashion Design. *See About Susan Scafidi*, COUNTERFEIT CHIC, http://about.counterfeitchic.com (last visited Jan. 1, 2011).

56. *See* Scafidi, *supra* note 49, at 117.

57. *Id.* Because the United States did not recognize copyrights for foreign authors, it was generally a haven for infringement. *Id.* at 118.

58. *See infra* Part II.B.1–4.

intellectual property law, including copyright and the useful articles doctrine, trademark and trade dress protection, and finally, patent law.

1. Intellectual Property under the Constitution

The United States Constitution gives Congress the power "[t]o promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries."[59] This constitutional guarantee is the heart of United States intellectual property protection in the form of copyright and patent law. By contrast, trademark protection arose from the common law and was enacted by Congress with the powers granted by the Commerce Clause.[60]

2. Copyright Law

Copyright protection applies to "original works of authorship fixed in any tangible medium of expression."[61] To receive protection under the Copyright Act, a work must be original.[62] This test is not stringent; essentially, it requires only a "modicum of creativity" and an independent expression.[63] Under Chapter 13, a design is original if it "provides a distinguishable variation over prior work pertaining to similar articles which is more than merely trivial and has not been copied from another source."[64] Copyright protection attaches at the moment of creation. However, in order to be eligible for certain infringement remedies and to file suit, the author must register the work with the Copyright Office.[65]

The Copyright Act grants the author the exclusive right to reproduce, distribute, perform, or display the protected work, and to create derivative works from the protected work.[66] Derivative works, whether created by the original author or by others, also qualify for protection, but this protection "does not extend to any part of the work in which such material has been used unlawfully."[67] Any such protection for a derivative work extends only

---

59. U.S. CONST. art. I, § 8, cl. 8.

60. *See* ROBERT P. MERGES, ET AL., INTELLECTUAL PROPERTY IN THE NEW TECHNOLOGICAL AGE 634 (rev. 4th ed. 2007).

61. 17 U.S.C. § 102(a) (2006).

62. Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 346 (1991).

63. *See id.*

64. 17 U.S.C. § 1301(b)(1) (2006). This language currently defines originality for vessel hull designs but would also apply to fashion designs under the DPPA and the IDPPPA. *Id.*; *see* S. 3728, 111th Cong. (2010); H.R. 2196, 111th Cong. (2009).

65. *See* 17 U.S.C. §§ 102(a), 412 (2006).

66. *See id.* § 106.

67. *Id.* § 103(a).

to the original work created by the later author.[68] Infringement occurs any time one of the author's exclusive rights is violated.[69]

Fashion designs have traditionally been denied copyright protection because they have been considered useful articles,[70] those having "an intrinsic utilitarian function that is not merely to portray the appearance of the article or to convey information."[71] Protection is not available for a "pictorial, graphic, [or] sculptural"[72] work—including fashion design—when the original aesthetic features cannot "be identified separately from, and are [not] capable of existing independently of, the utilitarian aspects of the article."[73] However, "an original design of a useful article which makes the article attractive or distinctive in appearance to the purchasing or using public may secure [*sui generis*] protection" under Chapter 13 of the Copyright Act,[74] which provides federal protection for some types of useful articles.[75] Currently, vessel hull designs that are "attractive or distinctive" are subject to protection under this Chapter,[76] but such designs enjoy only a ten-year period of protection.[77]

For other works, the key to determining protection for useful articles is separability. Protection is afforded when the design elements can be separated from the functional elements either physically or conceptually.[78]

Physical separability requires an article's artistic features to be independent of its functional aspects.[79] In 1954, the Supreme Court addressed the physical separability doctrine in *Mazer v. Stein*.[80] The case involved "statuettes of male and female dancing figures made of semivitreous [sic] china" used as a base of a lamp.[81] The issue before the Court was whether Stein could copyright the statuettes even though they

---

68. *See id.* § 103(b).

69. *See id.* § 501(a).

70. *See* Black, *supra* note 30, at 507.

71. 17 U.S.C. § 101 (2006). (defining "useful articles"). Under the statute, "[a]n article that is normally a part of a useful article is considered a 'useful article.'" *Id.*

72. *Id.*

73. *Id.* (defining "[p]ictorial, graphic, and sculptural works").

74. *Id.* § 1301(a)(1).

75. *See* 17 U.S.C. §§ 1301–1332(a) (2006).

76. *Id.* at § 1301.

77. *See id.* § 1305(a). This period is considerably longer than the three-year period that the DPPA and IDPPPA both propose. *See* S. 3728, 111th Cong. § 2(d) (2010); H.R. 2196, 111th Cong. § 2(d) (2009); *see also infra* Part IV.B.4.

78. *See* 17 U.S.C. § 101. Understanding both separability and its relationship to fashion design will aid in understanding the potential application of the DPPA and the IDPPPA because designs that these bills will protect still have to meet the utility test. *See id.* § 1302(4) (2006); S. 3728, 111th Cong. § 2(a)(2), (c) (2010); H.R. 2196, 111th Cong. § 2(c) (2009).

79. *See* 17 U.S.C. § 1302(4) (2006).

80. *See* Mazer v. Stein, 347 U.S. 201 (1954).

81. *Id.* at 202.

were part of a useful article.[82] The Court found that, because the statuettes were works of art that were physically separable from the lamp, they were entitled to copyright protection even though the creator intended to reproduce the work as part of a utilitarian object.[83] The Court added that Stein could prevent copies of the statuettes from being used in other lamps, but could not prohibit the general use of other statuettes as lamp bases.[84]

Similarly, in *Animal Fair, Inc. v. AMFESCO Industries, Inc.*, the United States District Court for the District of Minnesota upheld the copyright of a novelty slipper in the shape of a bear paw.[85] The slipper was "stuffed with foam and fiber," had a "tan sueded sole," and was "covered with a brown fur-like material."[86] The court found that "[v]irtually all of the design aspects of plaintiff's slipper 'can be identified separately from, and . . . exist . . . independently of, the utilitarian aspects of the [slipper].'"[87]

By contrast, in *Whimsicality, Inc. v. Rubie's Costume Co.*, the Second Circuit found that "Pumpkin, Bee, Penguin, Spider, Hippo Ballerina and Tyrannosaurus Rex" costumes were "soft sculptures" not subject to protection.[88] The court distinguished this case from *Animal Fair* on two grounds.[89] First, it found that slippers, unlike costumes, have "firm forms;" second, it found the application for copyright was misleading.[90] However, because the costumes' copyrights were denied based on their status as soft sculptures and not costumes, the court did not analyze any utilitarian aspects of the costumes.[91]

Conceptual separability exists "where design elements can be identified as reflecting the designer's artistic judgment exercised independently of functional influences."[92] A seminal case addressing conceptual separability in fashion design is *Kieselstein-Cord v. Accessories by Pearl, Inc.*, which addressed the issue "on [the] razor's edge of copyright law."[93] Designer Barry Kieselstein-Cord designed, manufactured, and sold fashion accessories, including two belt buckles called the "Winchester" and the "Vaquero."[94] Both buckles were art nouveau-inspired and cast in gold and silver "with rounded corners, a sculpted surface, . . . a rectangular cut-out at

82. *See id.* at 204–05.

83. *See id.* at 218; *see also* Judith S. Roth, et al., *Copyright Protection and Fashion Design*, 967 PLI/PAT 1081, 1087 (2009).

84. *See Mazer*, 347 U.S. at 218.

85. Animal Fair, Inc. v. AMFESCO Indus., 620 F. Supp. 175, 188 (D. Minn. 1985).

86. *Id.* at 178.

87. *Id.* at 187 (quoting 17 U.S.C. § 101 (2006)) (alteration in original).

88. Whimsicality, Inc. v. Rubie's Costume Co., 891 F.2d 452, 454, 457 (2d Cir. 1989).

89. *See id.* at 456.

90. *See id.*

91. *See id.* at 457.

92. Brandir Int'l, Inc. v. Cascade Pac. Lumber Co., 834 F.2d 1142, 1145 (2d Cir. 1987).

93. Kieselstein-Cord v. Accessories by Pearl, Inc., 632 F.2d 989, 990 (2d Cir. 1980).

94. *See id.*

one end for the belt attachment,"[95] and "several surface levels."[96] Kieselstein-Cord registered the belts with the Copyright Office, describing them as sculptures.[97] The Winchester buckle, which Kieselstein-Cord described as a "correlation between the art nouveau period and the butt of an antique Winchester rifle," sold 4,000 copies between 1976 and 1980 for between $295 and $6,000 each.[98] Accessories by Pearl began selling line-for-line copies of the buckles cast from "common metals" instead of gold and silver, referring to them as "'Barry K Copy,' 'BK copy,' and even 'Barry Kieselstein Knock-off.'"[99] Kieselstein-Cord sued Accessories by Pearl for infringement.[100]

The question before the Second Circuit was whether the buckles were eligible for copyright protection.[101] Accessories by Pearl argued that they could only qualify for protection if the buckles' sculptural aspects "[could] be identified separately from, and [were] capable of existing independently of, [their] utilitarian aspects."[102] The court, ruling in Kieselstein-Cord's favor, found that the buckles' sculptural elements were "conceptually separable" from their function, noting that some had worn the buckles in areas other than the belt, and that they "[had risen] to the level of creative art."[103]

In contrast, the court found that anatomically correct mannequins were not entitled to protection in *Carol Barnhart Inc. v. Economy Cover Corp.*[104] Carol Barnhart Inc. created polystyrene human torsos used to display clothing in retail stores.[105] Carol Barnhart brought suit against Economy Cover claiming it infringed the company's copyright "by offering for sale display forms copied from four original 'sculptural forms' to which [Carol] Barnhart owned the copyright."[106] Carol Barnhart argued that the mannequins "represent[ed] a concrete expression of a particular idea, e.g., the idea of a woman's blouse, and that the form involved, a human torso, is

---

95. *Id.* at 990 (internal quotations omitted).

96. *Id.*

97. *See id.* The Vaquero buckle was officially registered with the Copyright Office as jewelry, but Kieselstein-Cord's contribution "was listed on the certificate as 'original sculpture and design.'" *Id.* The registration for the Winchester buckle "specifically describe[d] the nature of the work as "sculpture." *Id.* at 991. Thus, although both buckles were not officially registered with the Copyright Office as sculptures, it is accurate to say that Kieselstein-Cord described them as such.

98. *Id.* at 991.

99. *Id.*

100. *See Kieselstein-Cord*, 632 F.2d 989.

101. *See id.* at 991.

102. *Id.* at 991–92.

103. *Id.* at 993–94.

104. *See* 73 F.2d 411 (2d. Cir. 1985).

105. *See id.* at 412.

106. *Id.*

traditionally copyrightable."[107] The court held that the forms were "utilitarian articles not containing separable works of art" and therefore, not subject to protection.[108]

Despite clothing's utilitarian nature, fashion designs have room to pass both the physical and conceptual separability tests. Many fashions today, such as Zac Posen's umbrella-sleeve blouse[109] or Hussein Chalayan's bubble dress,[110] are not dictated solely by the human form. Even if a design is more conventional in form than these extraordinary articles, its decorative aspects are not utilitarian in nature and therefore should receive protection. Courts have already laid the groundwork to protect several components of fashion design.[111] For example, both "design[s] printed upon dress fabric"[112] and silk-screen painting applied to ladies' blouses have been found to qualify for copyright protection, even when the "designs involve commonplace subject matters."[113] Despite these encouraging decisions, the courts must go further to protect fashion designs.

3. Trademark & Trade Dress Law

Like copyright, trademark and trade dress do protect some facets of fashion design, but they are insufficient to protect designers from the types of copying that they may face.

a. Trademarks

Trademarks are used to identify goods and services and distinguish them from others' products in the marketplace.[114] A trademark is a "word, name, symbol, or device, or any combination thereof" used in commerce.[115] Even colors, sounds, and other unregistered marks can be protected if they are distinctive or acquire secondary meaning.[116] Secondary meaning arises

---

107. *Id.* at 418.

108. *Id.* at 412.

109. *See H.R 5055 Hearing*, *supra* note 8, at 80 (statement of Susan Scafidi, professor at Fordham Law School, commenting on Zac Posen's umbrella-sleeved blouse).

110. *See* Amy Odell, *Does It Matter If Lady Gaga's Bubble Dress Is a Hussein Chalayan Knockoff?*, N.Y. MAG. (Mar. 23, 2009, 4:45 PM), http://nymag.com/daily/fashion/2009/03/does_it_matter_if_lady_gagas_b.html (discussing Hussein Chalayan's bubble dress being knocked off by Lady Gaga's designers).

111. *See* Scarves by Vera, Inc. v. United Merchs. & Mfrs., 173 F. Supp. 625, 627 (S.D.N.Y. 1959); Peter Pan Fabrics, Inc. v. Brenda Fabrics, Inc., 169 F. Supp. 142, 143 (S.D.N.Y. 1959).

112. *Peter Pan Fabrics*, 169 F. Supp. at 143.

113. *Scarves by Vera, Inc.*, 173 F. Supp. at 627.

114. *See* 15 U.S.C. § 1127 (2006).

115. *Id.*; *see id.* § 1051(a)(3)(C).

116. *See* SIEGRUN D. KANE, KANE ON TRADEMARK LAW §§ 2:10, 2:12 (5th ed. 2009).

when a consumer associates a mark or product with a single source.[117] This takes time to develop and usually as a result of repeated sales or advertising.[118]

A "spurious" mark "that is identical with, or substantially indistinguishable from," a registered trademark or service mark and that is used without permission is considered to be counterfeit.[119] Under current law, a mark infringes only when it is identical to, confusingly similar to, or dilutive of a famous trademark.[120] Famous trademarks include Chanel's interlocking, back-to-back C's[121] and Louis Vuitton's entwined "LV" Toile Monogram.[122] For example, in 2006 the Second Circuit found that Louis Vuitton's "Multicolore" mark featured on the designer's Murakami handbags was entitled to trademark protection.[123] The bags featured the Toile Monogram in thirty-three bright colors on a white or black background.[124] The court held that the use of the colors, combined with the mark, was "original in the handbag market and inherently distinctive."[125]

While trademark protection prohibits a person or company from putting another's name on their product, it does not necessarily prohibit them from putting their own name on a product that has "obvious similarities" to another's product as long as those similarities are not "confusingly similar."[126] In *Louis Vuitton Malletier v. Dooney & Bourke, Inc.*, the court found "obvious similarities" between Louis Vuitton's Murakami handbags and Dooney and Bourke's It-Bag.[127] The It-Bag "featured the [Dooney & Bourke] monogram in an array of bright colors set against a white background" and was made of fabric similar to the Murakami handbags.[128] However, regardless of the bags' similarities, they were not found to be

---

117. *Id.* at § 2:1.2[A].

118. *Id.*

119. 18 U.S.C. § 2320(e)(1)(A)(ii) (2006).

120. *See* 15 U.S.C. § 1125(c) (2006); 18 U.S.C. § 2320(e)(1)(A)(iv) (2006).

121. *See* Chanel, Inc. v. French, No. 05-61838-CIV, 2006 WL 3826780, at *1 (S.D. Fla. Dec. 27, 2006); *see also* CC, Registration No. 3,025,936.

122. *See* Louis Vuitton Malletier v. Dooney & Bourke, Inc., 454 F.3d 108, 116 (2d Cir. 2006); *see also* LV, Registration No. 2,909,002.

123. *See Louis Vuitton Malletier*, 454 F.3d at 116.

124. *See id.* at 112.

125. *Id.* at 116.

126. *See id.* at 117 (citing Louis Vuitton Malletier v. Burlington Coat Factory Warehouse Corp., No. 04 Civ. 2644, 2004 WL 1161167, at *8 (S.D.N.Y. May 24, 2004)) (discussing the district court's determination).

127. *See id.* The defendant, along with *Teen Vogue* magazine, used a team of teenage girls to develop the It-Bag. *Id.* at 113. The girls were photographed looking at the Louis Vuitton's monogrammed, multi-colored Murakami handbags through a store window and in a factory looking at swatches of the fabric used on the bags. *Id.*

128. *Id.*

confusingly similar.[129] *Louis Vuitton* is hence an excellent example of how trademark law alone does not protect all aspects of original designs.[130]

b. Trade Dress

In fashion, trade dress is "the total image of a good as defined by its overall composition and design, including size, shape, color, texture, and graphics."[131] There are two types of trade dress: (1) a product's packaging and labeling, and (2) its overall appearance and configuration.[132] In other words, trade dress "'includes the design and appearance of the product . . . and all elements making up the total visual image by which the product is presented to customers.'"[133]

Under the trade dress doctrine, a product can receive protection only if it is non-functional and either is inherently distinctive or has secondary meaning.[134] These requirements make obtaining trade dress protection for fashion design nearly impossible.[135] In *Wal-Mart Stores v. Samara Brothers, Inc.*, the Supreme Court determined that product design, as opposed to product packaging, cannot be inherently distinctive and therefore must have achieved secondary meaning to be protected.[136] Samara Brothers designed and manufactured children's clothing for retailers like JCPenney.[137] The designs were "a line of spring [and] summer one-piece seersucker outfits decorated with appliqués of hearts, flowers, fruits, and the like."[138] Wal-Mart sent photographs of the Samara Brothers designs to one of its suppliers, which then copied sixteen of the garments with only minor changes.[139] The low-cost knockoff Wal-Mart designs were so similar to the look and feel of the higher end Samara Brothers designs that a

129. *See id.* at 117. The appellate court remanded the issue of likelihood of confusion to the trial court, yet the trial court never directly addressed the issue. *See* Louis Vuitton Malletier v. Dooney & Bourke, Inc. 500 F. Supp. 2d 276 (S.D.N.Y. 2007).

130. The current trademark infringement standards make it more difficult to prove infringement than the "closely and substantially similar" standard provided for in the DPPA, see H.R. 2196, 111th Cong. § 2(e)(2) (2009), but is similar in strength to the "substantially identical" standard proposed by the IDPPPA, see S. 3728, 111th Cong. § 2(e)(2) (2010). The proposed bills may be able to protect against designers who have intentionally copied an original style, yet do not quite reach the level of being confusingly similar. *See infra* Part III.B.6.b.

131. Coach Leatherware Co. v. Ann Taylor, Inc., 933 F.2d 162, 168 (2d Cir. 1991).

132. *See* KANE, *supra* note 116, § 3:1.

133. Knitwaves, Inc. v. Lollytogs Ltd., 71 F.3d 996, 1005 (2d Cir. 1995) (quoting Jeffrey Milstein, Inc. v. Greger, Lawlor, Roth, Inc., 58 F.3d 27, 31 (2d Cir. 1995)).

134. *See* KANE, *supra* note 116, at § 3:2.

135. *See* Beltrametti, *supra* note 29, at 154–55.

136. *See* Wal-Mart Stores v. Samara Bros., 529 U.S. 205, 216 (2000).

137. *See id.* at 207.

138. *Id.*

139. *See id.* at 207–08.

JCPenney representative mistook them for the originals.[140] Although many of the designs had copyrighted elements, the Court held that there was no infringement under trade dress because the designs did not have secondary meaning.[141] Justice Scalia, writing for the majority, explained that in the Court's view, "consumer predisposition to equate the [product design] with the source does not exist."[142]

Apparently, Justice Scalia does not know any fashionistas. It does not take a trained eye to recognize immediately Hervé Léger's bandage dresses[143] or Alexander McQueen's "Armadillo" shoes.[144] Unfortunately, despite the fact that both are undeniably identifiable, they are barred from automatic protection.[145] The problem stems from the fact that it takes time to establish secondary meaning, and in the meantime, designers are vulnerable to copyists and may incur large attorneys' fees while trying to protect their work.[146] The presence of copies on the market makes it even harder to establish secondary meaning—it is difficult for a company to establish itself as the sole source of a good if there are competitors selling exactly the same product.[147] For example, even though Christian Louboutin has been using red-lacquered soles on his shoes since 1992, he did not receive trademark protection for their use until January 1, 2008,[148] and he was able to do so then only after multiple battles with other designers who used his signature look.[149]

---

140. *Id.* at 208.

141. *See id.* at 216.

142. *Id.* at 213. The Court found "that design . . . is not inherently distinctive." *Id.* at 212.

143. *See* HERVE LEGER, http://www.herveleger.com (last visited Jan. 1, 2011).

144. *See* Hilary Moss, *Three Models Cut from Alexander McQueen Show After Refusing to Wear Armadillo Shoes*, HUFFINGTON POST (Dec. 22, 2009, 9:32 AM), http://www.huffingtonpost.com/2009/12/22/three-models-cut-from-ale_n_400440.html

145. Because a product design must develop secondary meaning, immediate protection is impossible. *See Wal-Mart Stores*, 529 U.S. at 216.

146. An element of trade dress, such as color, would not provide immediate protection but could become associated with a brand over time. *See* Qualitex Co. v. Jacobson Prods. Co., 514 U.S. 159, 163 (1995).

147. *See generally* KANE, *supra* note 116.

148. *See* Susan Scafidi, *Happy New Year to Christian Louboutin!*, COUNTERFEIT CHIC (Jan. 13, 2008, 7:16 PM), http://www.counterfeitchic.com/2008/01/happy_new_year_to_chris tian_lo_1.php.

149. *See* Susan Scafidi, *Kors Light*, COUNTERFEIT CHIC (July 6, 2009), http://counterfeitchic.com/2009/07/kors-light.html (discussing Michael Kors's use of the red soles); Susan Scafidi, *Piracy by Prada?!*, COUNTERFEIT CHIC (July 4, 2007, 07:18 PM), http://www.counterfeitchic.com/2007/07/miu_miu_too.php (discussing Prada's Miu Miu line's use of red soles on similar looking shoes); Susan Scafidi, *Seeing Red*, COUNTERFEIT CHIC (March 9, 2007, 10:27 AM), http://www.counterfeitchic.com/2007/03/seeing_red.php (discussing that the red-soled shoes were copied by Oh...Deer!).

Another problem lies with the fact that, to obtain protection under the trade dress doctrine, a designer must "establish the nonfunctionality of the design feature," and clothing is generally considered to have function.[150] Courts use two tests to determinine whether a fashion design is functional: the "alternative designs" and "aesthetic functionality" tests.[151] The former examines whether the element of the trade dress "is essential to the use or purpose of the article or if it affects the cost or quality of the article"[152] while the latter examines whether the "exclusive use of [an element of the trade dress] would put competitors at a significant non-reputation-related disadvantage."[153] If a design's sole purpose is to be aesthetically pleasing, it is barred from receiving trade dress protection.[154] However, that is the primary thrust of fashion design: its purpose is to be beautiful. While trademark and trade dress law provide limited protection to certain areas of fashion designs, they do not adequately safeguard the rights of designers.

4. Patent Law

Like copyright and trademark law, patent law does not offer sufficient protection for fashion designs. For an item to be eligible for patent protection, it must be a patentable subject matter, useful, novel, non-obvious, and properly disclosed to the Patent Office.[155] A design may receive design patent protection when it is a "new, original and ornamental design for an article of manufacture."[156] Essentially, a design patent covers the decorative aspect of an article instead of its utility.[157] In fact, if an article's design is solely determined by its function, the article is not eligible for a design patent.[158] If a designer obtains a patent on a design, the designer has the right to prevent others from using it for fourteen years.[159]

---

150. *Wal-Mart Stores*, 529 U.S. at 214 (citing 15 U.S.C. § 1125(a)(3) (1994 ed., Supp. V)).

151. *See* Roth et al., *supra* note 83, at 1087 (citing Knitwaves, Inc. v. Lollytogs Ltd., 71 F.3d 996, 1006 (2d Cir. 1995)).

152. TrafFix Devices, Inc. v. Marketing Displays, Inc., 532 U.S. 23, 32 (2001) (quoting Qualitex Co. v. Jacobson Products Co., 514 U.S. 159, 165 (1995)) (internal citation omitted).

153. *Id.* (quoting *Qualitex Co.*, 514 U.S. at 165) (internal citation omitted). Essentially, the court determines whether the feature is used to be aesthetically pleasing or instead to designate the source. Knitwaves, Inc. v. Lollytogs Ltd., 71 F.3d 996, 1006 (2d Cir. 1995) (citing LeSportsac, Inc. v. K-mart Corp., 754 F.2d 71, 78 (2d Cir. 1985)).

154. *See id.*

155. *See* 35 U.S.C. §§ 101–103, 112 (2006).

156. *Id.* § 171.

157. *See* MERGES ET AL., *supra* note 60, at 371.

158. *See* Anne Theodore Briggs, *Hung Out To Dry: Clothing Design Protection Pitfalls in United States Law*, 24 HASTINGS COMM. & ENT. L.J. 169, 176 (2002) (citing DONALD S. CHISUM, CHISUM ON PATENTS § 1.04(2) (1997)).

159. 35 U.S.C. § 173 (2006).

The main hurdle for fashion designs to obtain design patents is the nonobviousness requirement.[160] Nonobviousness requires more than a trivial advance over what has been done before from the perspective of someone skilled in the relevant art.[161] To determine what is obvious, a court would consider all pertinent prior existing designs, alone or in combination.[162] If a subsequent design resulted from a combination of the prior designs that required no more "ingenuity and skill" than that possessed by a designer within that genre, it would not receive protection.[163]

Furthermore, the cost and lengthy process of patent registration precludes its usefulness in solving the problems designers face.[164] The cost of filing and receiving a design patent is $740 per application, plus attorney fees.[165] This becomes a hefty fee when a designer has multiple pieces in each season's collection, and it is likely beyond the means of designers who are just beginning their careers.[166] Moreover, it takes an average of twenty-five months to receive a design patent.[167] This is entirely too long to be effective in the fashion industry, which moves at a blinding pace.[168] By the time designers learn whether their designs have received protection, they will have already moved on to creating new designs.[169]

C. Prior Congressional Consideration & Attempts to Obtain Protection for Fashion Design

The fight to gain recognition for original fashion designs is not a new one. The DPPA and the IDPPPA mark the ninety-second and ninety-third attempts, respectively, since 1914 to obtain legislative protection for

---

160. *See generally* Briggs, *supra* note 158.

161. ROBERT PATRICK MERGES & JOHN FITZGERALD DUFFY, PATENT LAW AND POLICY: CASES AND MATERIALS 96 (4th ed. 2007) (citing 35 U.S.C § 103 (2006)).

162. *Id.* at 663.

163. *See* Graham v. John Deere Co. of Kansas City, 383 U.S. 1, 8 (1966) (quoting Hotchkiss v. Greenwood, 52 U.S. (111 How.) 248, 267 (1851)).

164. *See* 35 U.S.C. § 41(a)(3) (2006); Julie P. Tsai, Comment, *Fashioning Protection: A Note on the Protection of Fashion Designs in the United States*, 9 LEWIS & CLARK L. REV. 447, 457 (2005).

165. *See* 35 U.S.C. § 41(a)(3) (2006) (stating that each design application costs $310 and each issued design patent costs $430).

166. *See Average Income for Fashion Designers*, FASHION DESIGNER SALARY, http://www.fashiondesignersalarydata.com/fashion-designer-salary.html (last visited Jan. 1, 2011) ("[E]ntry level fashion designers can earn less than $30,000 a year.").

167. *See Questions and Answers–USPTO*, U.S. PATENT & TRADEMARK OFFICE, http://www.uspto.gov/main/faq/ (last visited Jan. 1, 2011); *see also* Black, *supra* note 30, at 507.

168. *See* Black, *supra* note 30, at 507.

169. *See id.*

fashion designs.[170] This section will discuss some of the historical attempts to gain design protection and the recent bills introduced in Congress.

1. The Design Registration League

In 1914, not long after the Copyright Act was first adopted in 1909, the Design Registration League (the "League") drafted a bill that would provide registration for "'any design, new and original, as embodied in or applied to any manufactured product of an art or trade.'"[171] The League, which included pocket watch, lace, and embroidery manufacturers, was "an organized group of companies which believed that they 'were unable to obtain satisfactory protection for their original design work' under the then present statutes."[172] Among their complaints was that "[t]he designer is essentially an artist and should be protected as other artists are protected."[173] The House Committee was concerned that the bill was too broad and thus requested a redrafted version.[174] Similar versions followed in 1916 and 1917, and though they were ultimately unsuccessful and neither bill made it to a floor vote, the 1917 version did "[come] out of conference committee with a unanimous vote."[175]

2. Fashion Originators' Guild of America

In 1932, because of the lack of formal intellectual property protection in the United States, the Fashion Originators' Guild of America ("the Guild") formed to take private action to prevent style pirates from selling copies of their goods, a practice the Guild called "unethical and immoral."[176] The Guild was a group of designers and manufacturers who

---

170. *See* Lauren Estrin et al., *In Vogue: IP protection for Fashion Design*, COPYRIGHT WORLD, Apr. 2007, at 21, 24, *available at* http://www.kilpatrickstockton.com/~/media/files/articles/invogueipprotectionforfashiondesign.ashx (noting that H.R. 5055 was the "89th failed attempt since 1914 to adopt US copyright legislation to protect fashion designs"). Since then, S. 1957 and H.R. 2033 also failed to pass; H.R. 2196 and S. 3728 are now pending. *See* S. 3728, 111th Cong. (2010); H.R. 2196, 111th Cong. (2009); S. 1957, 110th Cong. (2007); H.R. 2033, 110th Cong. (2007).

171. David Goldenberg, *The Long and Winding Road, A History of the Fight over Industrial Design Protection in the United States*, 45 J. COPYRIGHT SOC'Y U.S.A. 21, 27 (1997) (quoting *Registration of Designs: Hearings on H.R. 6458 and H.R. 13618 Before the H. Comm. on Patents*, 64th Cong. 3 (1916)).

172. *Id.* at 28 (quoting *Registration of Designs: Hearings on H.R. 6458 and H.R. 13618 Before the H. Comm. on Patents*, 64th Cong. 13 (1916) (statement of E.W. Bradford, General Counsel)).

173. *Id.* (citation omitted).

174. *See id.* at 30–31. (citation omitted).

175. *See id.* at 31.

176. *See* Fashion Originators Guild of Am. v. Fed. Trade Comm'n, 312 U.S. 457, 461 (1941).

sold moderate-to-high-priced original designs to retailers.[177] As a method of self-help, the Guild established a "Piracy Committee" to determine whether the members' designs were indeed original and whether retailers were selling knockoffs of those designs.[178] The Guild and its members then refused "to sell any dresses to retailers who purchase[d], or order[ed] to be manufactured, dresses which the Guild [found embodied] copies of its designs."[179]

In addition to regulating the designer-to-retailer stream of commerce, the Guild had a division called the "'Textile Affiliates or Associates' whose members register[ed original] textile designs with the 'National Federation of Textiles, Inc.'"[180] The Guild's dressmakers had to agree not to buy "'unregistered' fabrics," and the textile members had to agree to sell only to the dressmaker members.[181] If the Guild discovered that its members bought from or sold to nonmembers or sold to red-carded retailers (those who were caught selling copies), it subjected the violators to fines.[182] The Guild claimed that the sanctions "were necessary to protect the industry as a whole from 'demoralization' and the 'property' of its members from appropriation."[183]

The Guild eventually received a cease-and-desist order from the Federal Trade Commission because the Commission concluded that the Guild's policies were "an unfair method of competition" and "[ran] counter to the public policy declared in the Sherman and Clayton [Antitrust] Acts."[184] In 1941, the United States Supreme Court agreed with the Commission, finding that the Guild's purpose "was the intentional destruction of one type of manufacture and sale which competed with Guild members" and that the Guild's operations were an "unlawful combination."[185] Over the next sixty-five years, regular attempts were made to pass laws to protect fashion design, yet none were successful.[186] However, a new push began in 2006 with the introduction of H.R. 5055.[187]

---

177. *See Fashion Originators Guild*, 114 F.2d at 81–82. "In 1936, [the Guild's manufacturer members] sold in the United States more than 38% of all women's garments wholesaling at $6.75 and up, and more than 60% of those at $10.75 and above." *Fashion Originators Guild*, 312 U.S. at 462.

178. *Fashion Originators Guild*, 114 F.2d at 82. Members registered designs with the Guild (much like the DPPA requires) for them to determine if the designs were originals. *Id.*

179. *Id.*

180. *Id.*

181. *Id.*

182. *See Fashion Originators Guild*, 312 U.S. at 462–63.

183. *Fashion Originators Guild*, 114 F.2d at 82.

184. *Fashion Originators Guild*, 312 U.S. at 460, 463.

185. *Id.* at 467–68.

186. *See* Estrin, *supra* note 170, at 24.

187. *See* H.R. 5055, 109th Cong. (2006).

3. H.R. 5055: To Amend Title 17, United States Code, to Provide Protection for Fashion Design, 109th Congress

Virginia Representative Bob Goodlatte introduced H.R. 5055 on March 30, 2006.[188] It was the first in a series of fashion design protection bills introduced from 2006 to the present, and it has served as a foundation for later versions of similar bills.[189] H.R. 5055 would have provided protection only to qualifying fashion designs "that [were] made public by the designer or owner in the United States or a foreign country" within three months of "the date of the application for registration."[190] This meant that a design had to be registered with the Copyright Office to obtain protection.[191]

H.R. 5055 never made it out of the Committee.[192] On July 26, 2006, the House Judiciary Subcommittee on Courts, the Internet, and Intellectual Property heard testimony from fashion designers and law professors, both in support of and in opposition to the bill.[193] While maintaining an open mind, California Representative Howard Berman questioned the need for the legislation, asking, "Is a high level of protection necessary to promote innovation? Or does the lack of a high level of protection for fashion designs actually spur increased creativity in the fashion industry?"[194] In discussing his apprehensiveness, he stated that he needed more information, both on what the Copyright Act's "substantially similar" test[195] would include and whether the legislation would cover industry trends.[196]

---

188. *See id.*

189. *See generally* S. 3728, 111th Cong. (2010); H.R. 2196, 111th Cong. (2009); S. 1957, 110 Cong. (2007); H.R. 2033, 110th Cong. (2007); H.R. 5055, 109th Cong. (2006).

190. H.R. 5055, 109th Cong. § 1(b)(3), 1(e)(1) (2006). Unlike its successor, the DPPA, H.R. 5055 included neither a searchable database for the registered designs nor specific language that excluded from protection designs that merely reflected a trend. *See* H.R. 2196, 111th Cong. § 2(e)(3), (j) (2009); H.R. 5055, 109th Cong. (2006).

191. *See* Nicole Giambarrese, Comment, *The Look for Less: A Survey of Intellectual Property Protections in the Fashion Industry*, 26 TOURO L. REV. 243, 253–54 (2010).

192. The last action on H.R. 5055 was a subcommittee hearing on July 27, 2006. *See H.R. 5055: To Amend Title 17, United States Code, to Provide Protection for Fashion Design*, GOVTRACK.US, http://www.govtrack.us/congress/bill.xpd?bill=h109-5055 (last visited Jan. 1, 2011) [hereinafter *H.R. 5055 GovTrack*].

193. *See generally H.R. 5055 Hearing*, *supra* note 8. American designer Jeffrey Banks and Fordham Law School professor Susan Scafidi testified in favor of the bill; David Wolfe, creative director of The Doneger Group, and University of Virginia School of Law professor Christopher Sprigman testified in opposition. *Id.* at 8–87.

194. *Id.* at 2–3 (statement of Rep. Berman).

195. *See* 17 U.S.C. § 1309(e) (2006). H.R. 5055 would have incorporated the "substantially similar" test in Chapter 13 of the Copyright Act. H.R. 5055, 109th Cong. § 1(d) (2006).

196. *See H.R. 5055 Hearing*, *supra* note 8, at 3 (statement of Rep. Berman) ("In addition, I would like for the witnesses to describe what constitutes a design that is substantially similar. Is it an exact copy? Is it a mere inspiration of a current trend? And how

Trends also concerned David Wolfe, creative director of Doneger Creative Services, who testified in opposition to the legislation during the H.R. 5055 hearings, which prompted Representative Goodlatte to ask whether adding a provision to exclude trends would improve the bill.[197] Mr. Wolfe stated that a "trend" was difficult to define and that "the whole fashion concept is so ephemeral that trying to nail down specifics becomes impossible."[198] Moreover, in response to a similar question, another opponent stated that the bill would be better if it was "clearly limited only to those garments that are point-by-point copies of existing garments."[199]

4. H.R. 2033 & S. 1957: Design Piracy Prohibition Act, 110th Congress

On April 25, 2007, Massachusetts Representative Bill Delahunt introduced H.R. 2033, the first bill to be titled the Design Piracy Prohibition Act.[200] Although the proposed bill was identical to H.R. 5055, it picked up an additional seven cosponsors, for a total of fourteen.[201] At the subcommittee hearing on February 14, 2008, Representative Berman again voiced his concerns over the degree of similarity required for infringement.[202] Once again, the bill never made it out of the Committee.[203]

Senate Bill 1957, introduced on August 2, 2007, began to move toward compromise with opponents.[204] S. 1957 was the first to include a provision that limited the bill's scope as it pertained to the similarity of infringing articles.[205] It provided that a design would not be considered to be copied "if it is original and not *closely and substantially similar* in overall visual appearance to a protected design."[206] The bill was referred to the Senate Committee on the Judiciary, but hearings were never held.[207]

---

does one determine if it is something in between?").

197. *See id.* at 6 (statement of Rep. Goodlatte).

198. *Id.* (statement of David Wolfe).

199. *Id.* at 7 (statement of Christopher Sprigman, professor at the University of Virginia School of Law).

200. H.R. 2033, 110th Cong. (2007).

201. *See H.R. 2033: Design Piracy Prohibition Act*, GOVTRACK.US, http://www.govtrack.us/congress/bill.xpd?bill=h110-2033 (last visited Jan. 1, 2011) [hereinafter *H.R. 2033 GovTrack*]; *H.R. 5055 GovTrack*, *supra* note 192.

202. *See H.R. 2033 Hearing*, *supra* note 10, at 2 (statement of Rep. Berman) ("How similar must a design be before you can enforce the design right?").

203. *See H.R. 2033 GovTrack*, *supra* note 201. On May 4, 2007, H.R. 2033 was referred to the Subcommittee on Courts, the Internet, and Intellectual Property, which was the last action taken. *Id.*

204. *See* S. 1957, 110th Cong. (2007).

205. *See id.* § 2(d)(2).

206. *Id.* § 2(d)(2)(C) (emphasis added).

207. *See S. 1957: Design Piracy Prohibition Act*, GOVTRACK.US, http://www.govtrack.us/congress/bill.xpd?bill=s110-1957 (last visited Jan. 1, 2011) No actions were taken on the

5. H.R. 2196: Design Piracy Prohibition Act & S. 3728: Innovative Design Protection and Piracy Prevention Act, 111th Congress

On April 30, 2009, Representative Delahunt again introduced the DPPA as H.R. 2196.[208] The bill was referred to the House Committee on the Judiciary on the same day and progressed no further.[209] However, with twenty-three sponsors, the bill gained more support than previous versions.[210]

On August 5, 2010, Senator Charles Schumer introduced S. 3728, the IDPPPA.[211] On December 1, 2010, the Senate Committee on the Judiciary held an executive business meeting in which it voted unanimously for the bill to proceed to the Senate floor.[212] This is the furthest that any of the design bills has progressed since 2006.[213]

In the meeting, Senator Schumer stated that the bill had "broad bipartisan support"[214] and that it was the result of "a year of extensive negotiations between the major players in the industry."[215] However, Senator John Cornyn expressed concern that no hearings had been held to discuss the impact of the IDPPPA, but he did not oppose moving the bill to the floor.[216] On December 6, 2010 the IDPPPA was reported to the Senate with an amendment in the nature of a substitute.[217]

Congress adjourned December 22, 2010 before either bill came to a vote.[218] This means that in order for fashion design protection to become

---

bill after it was referred to the Committee on the Judiciary. *Id.*

208. *See* H.R. 2196, 111th Cong. (2009).

209. *See H.R. 2196: Design Piracy Prohibition Act*, GOVTRACK.US, http://www.govtrack.us/congress/bill.xpd?bill=h111-2196 (last visited Jan. 1, 2011) [hereinafter *H.R. 2196 GovTrack*].

210. *See id.*; *see also H.R. 2033 GovTrack*, *supra* note 201.

211. *See* S. 3728, 111th Cong. (2010).

212. Executive Business Meeting to Consider Pending Nominations and Legislation, JUDICIARY.SENATE.GOV,http://www.senate.gov/fplayers/CommPlayer/commFlashPlayer.cfm?fn=judiciary120110&st=xxx, at 50:20, (last visited Jan. 1, 2011) [hereinafter Executive Business Meeting]; Alison A. Nieder, *Proposed Design Piracy Law Moves Forward*, CALIFORNIA APPAREL NEWS, (Dec. 1, 2010), http://www.apparelnews.net/news/manufacturing/120110-Proposed-Design-Piracy-Law-Moves-Forward.

213. *See* S. 3728, 111th Cong. (2010); H.R. 2196, 111th Cong. (2009); S. 1957, 110th Cong. (2007); H.R. 2033, 110th Cong. (2007); H.R. 5055, 109th Cong. (2006),

214. Executive Business Meeting, *supra* note 212, at 50:39.

215. *Id.* at 51:42.

216. *Id.* at 53:30. Senator Cornyn also "circulated an amendment . . . aimed at minimizing [litigation] costs by imposing a loser pay rule" but did not offer the amendment during the meeting. *Id.* at 53:42.

217. Senator Schumer introduced a manager's amendment to this bill discussed *infra* Part III.B. *See* S. 3728 (as reported by S. Comm. On the Judiciary, Dec. 6, 1989), 111th Cong. (2010); Executive Business Meeting, *supra* note 212, at 52:10.

218. *See H.R. 2196 GovTrack*, *supra* note 209; *see S. 3728: Innovative Design*

law a new bill must be reintroduced by the 112th Congress. These bills will be discussed in greater detail in Part III.B.

III. ANALYSIS

A. Why Is Fashion Design Protection Necessary?

1. Fashion is Money

Fashion is a "global marketplace worth hundreds of billions of dollars."[219] The United States market alone accounts for $196 billion,[220] approximately $12 billion of which is from the sale of knockoffs.[221] The industry is not composed solely of the 22,000 fashion designers nationwide; it also includes over four million fabric manufacturers, retailers, models, seamstresses, sales persons, publicists, tailors, and various support staff like truck drivers.[222] According to Arie Kopelman, vice chairperson and past president of Chanel Inc., design piracy places these jobs at risk.[223] Creating a new fashion line is extremely expensive.[224] Not only does it require tens of thousands of dollars to start a clothing line, but a similar investment is required each season to develop original prints, designs, and samples.[225] Moreover, it costs between $50,000 and $1,000,000 to produce a fashion show to inform the press and the public that the designs even exist.[226] All of these events and costs occur before the season's first orders are even placed.[227]

Protection and Piracy Prevention Act, GOVTRACK.US, http://www.govtrack.us/congress/bill.xpd?bill=s111-3728 (last visited Jan. 1, 2011) [hereinafter *S. 3728 GovTrack*].

219. STOP FASHION PIRACY, *supra* note 9.

220. *See* Maggie Overfelt, *When Piracy is Legal*, CNN MONEY (Apr. 28, 2008, 9:20 AM), http://money.cnn.com/2008/04/23/smbusiness/whos_stealing_piracy_legal.fsb ("The Council of Fashion Designers of America (CDFA), a nonprofit trade group, estimates that knockoffs represent at least five percent of the nation's $196 billion apparel market.").

221. *See H.R. 2033 Hearing*, *supra* note 10, at 22 (statement of Narcisco Rodriguez, fashion designer).

222. *See* BUREAU OF LABOR STATISTICS, U.S. DEP'T OF LABOR, *Fashion Designers*, *in* OCCUPATIONAL OUTLOOK HANDBOOK1, 1–2 (2010–2011 ed. 2009), *available at* http://www.bls.gov/oco/pdf/ocos291.pdf (last modified Dec. 17, 2009); *Number of People Are Employed in the Fashion Industry*, GRABSTATS.COM, http://www.grabstats.com/statmain.asp?StatID=1018 (last visited Jan. 1, 2011); STOP FASHION PIRACY, *supra* note 9.

223. *See* STOP FASHION PIRACY, *supra* note 9 (statement of Arie Kopelman, vice chairperson & past president of Chanel Inc.).

224. *See H.R. 5055 Hearing*, *supra* note 8, at 11–12 (statement of Jeffrey Banks).

225. *See id.* at 12.

226. *See id.*

227. *See id.*

Unfortunately, after all the expense and hard work, the design "can be stolen before the applause has faded."[228] Making a marketable knockoff product used to take between several months and a year.[229] Now, technological advancements have made copying nearly instantaneous.[230] Often, the copied products are sold before the original design even reaches the stores.[231] For example, just days after Chelsea Clinton's wedding, famed knockoff artist Allen Schwartz of ABS announced that he would create a replica of the Vera Wang gown that Clinton wore down the aisle.[232] This process is particularly harmful to small designers because they cannot simply absorb their lost profits and move on.[233] Zac Posen, an American fashion designer, says that this design piracy can derail careers.[234] Larger American designers, like smaller, start-up designers, struggle to compete with overseas companies that produce line-for-line knockoffs using cheaper labor.[235] This may eventually lead to the loss of jobs if designers leave the United States for other Western countries that provide legal protection for fashion designs.[236]

Even if the industry has historically thrived with piracy, this success was due in large part to the time delay.[237] This is a species of the "first-mover advantage" often discussed in relation to the need for intellectual property protection.[238] With modern technology that delay has vanished, and with it, the advantages of the original designer as the first mover.[239] Even if we didn't need fashion design protection before, we do now.

2. Fashion is a Form of Art and Should Receive Protection Equivalent to the Traditional Fine Arts

The right to intellectual property ownership is crucial for encouraging innovation among our country's artists and scientists, and the exact same reasoning applies to fashion design. This highlights the need to amend the Copyright Act to include fashion design. As noted at the outset of this

---

228. *Id.*

229. *See id.*

230. *See id.* at 9.

231. *See id.*

232. *See* Eric Wilson, *No Wonder Chelsea Clinton Wanted Secrecy*, N.Y. TIMES, Aug. 5, 2010, at E4, *available at* http://www.nytimes.com/2010/08/05/fashion/05ROW.html.

233. *See* STOP FASHION PIRACY, *supra* note 9.

234. *See id.* (statement of Zac Posen, American fashion designer.).

235. *See id.* (statement of Arie Kopelman).

236. *See H.R. 5055 Hearing*, *supra* note 8, at 9 (statement of Jeffrey Banks) ("The [United States] is conspicuous in that unlike Europe and Japan, it does not protect fashion in its laws."); STOP FASHION PIRACY, *supra* note 9 (statement of Zac Posen).

237. *See H.R. 5055 Hearing*, *supra* note 8, at 12 (statement of Jeffrey Banks).

238. *See* Kal Raustiala & Christopher Sprigman, *The Piracy Paradox: Innovation and Intellectual Property in Fashion Design*, 92 VA. L. REV. 1687, 1159–62 (2006).

239. *See H.R. 5055 Hearing*, *supra* note 8, at 12 (statement of Jeffrey Banks).

comment, the Supreme Court stated in *Mazer* that "[i]t is clear Congress intended the scope of the copyright statute to include more than the traditional fine arts."[240] Copyright protections are broad and cover books, music, plays, paintings, sculptures, movies, architectural works, and even this Comment.[241] Though fashion design is often considered an "applied art"[242] instead of a traditional fine art, the development, visual impact, and skill required to create fashion designs are just as significant as traditionally protected arts and are worthy of equal protection. Fashion designers often spend years learning their craft, both in school and through internships.[243] But like any art, fashion designers must possess a natural talent to accompany the skills they have developed through study.[244] Designers need "an eye for color and detail, a sense of balance and proportion, and an appreciation for beauty."[245] Moreover, fashion designers must be able to put it all together with sewing skills and translate their ideas into functional patterns.[246]

Couture has been featured in art museums across the world.[247] The Victoria & Albert Museum's exhibition, *The Golden Age of Couture: Paris and London 1947–1957*, has been presented in London, Australia, Canada, Hong Kong, and even Nashville, Tennessee.[248] The exhibition has been so popular that its 2007 opening at the Victoria & Albert Museum broke the museum's attendance records.[249] In fact, the attendance for fashion exhibitions at museums has increased worldwide in recent years.[250] This growing acceptance of fashion designs as works of art demonstrates that it is time for intellectual property law in the United States to acknowledge their original and elegant contributions to society.

---

240. Mazer v. Stein, 347 U.S. 201, 213 (1954).

241. *See* 17 U.S.C. § 102(a) (2006).

242. *See* Black, *supra* note 30, at 521.

243. *See* BUREAU OF LABOR STATISTICS, U.S. DEP'T OF LABOR, *supra* note 222 (explaining that designers spend two to four years studying the elements of design and often learn the craft through internships).

244. *See id.*

245. *Id.*

246. *See id.*

247. *See The Golden Age of Couture*, THE FRIST CENTER FOR THE VISUAL ARTS, http://www.fristcenter.org/site/exhibitions/exhibitiondetail.aspx?cid=795 (last visited Jan. 1, 2011).

248. *See id.*

249. *See id.*

250. *See* Lottie Johansson, *Grace and Style*, *available at* http://webartacademy.com /grace-and-style, (last visited Jan. 1, 2011).

3. Piracy Affects Designers, Large and Small

Both world-famous and independent designers alike have lost business due to design piracy.[251] The following stories illustrate how design piracy has directly affected designers' income.[252] Additionally, they demonstrate that the availability of low-quality or low-priced knockoffs does affect the profits of higher-priced designers.[253] Each of these designers was successfully selling its designs but experienced returns, decreased sales, or canceled orders once knockoff products appeared in the market.[254]

Nicole Dreyfuss, a New York handbag designer, created $130 handmade cable-knit clutches that featured "wooden dowels and [were] bedecked in preppy-looking ribbons."[255] While browsing the Internet, she discovered someone urging an admirer of her work not to purchase a clutch because Abercrombie & Fitch was selling "the same bag" for only $30.[256] Ms. Dreyfuss indicated that competing with such a large company was impossible.[257] To do so, she would have to sell her product for $15, a price that would not even cover the supplies.[258]

Emerging scarf designer Elle Sakellis recently discovered that her original "evil eye" scarves had been knocked off when one of her retailers did not renew its order.[259] Ms. Sakellis's scarves were selling well until the cheaper product hit the market.[260] The store's owner admitted that the availability of a cheaper version was the reason for his not renewing its order, stating that "'[i]t's like, we carried the higher price, . . . and now we're carrying the diffusion line.'"[261] The scarves were identical in style and color, but her $190 scarves were made of silk chiffon, while the $10 knockoffs were made of polyester.[262]

Small designers are not the only ones losing money.[263] Larger designers are not only seeing customers turning to cheaper designs looking for a deal,

---

251. *See* Christina Binkley, *The Problem with Being a Trendsetter*, WALL ST. J., Apr. 29, 2010, at D8, *available at* http://online.wsj.com/article/NA_WSJ_PUB:SB10001424052748704423504575212201552288996.html; Kover, *supra* note 41; Overfelt, *supra* note 220.

252. *See* Binkley, *supra* note 251; Kover, *supra* note 41; Overfelt, *supra* note 220.

253. *See* Binkley, *supra* note 251; Kover, *supra* note 41; Overfelt, *supra* note 220.

254. *See* Binkley, *supra* note 251; Kover, *supra* note 41; Overfelt, *supra* note 220.

255. Kover, *supra* note 41.

256. *Id.*

257. *See id.*

258. *Id.* Ms. Dreyfuss's attorney sent a cease-and-desist letter to the large chain. *Id.* To her surprise, Abercrombie & Fitch complied and agreed to pull the purses off of its shelves. *Id.*

259. *See* Binkley, *supra* note 251.

260. *See id.*

261. *Id.*

262. *See id.*

263. *See* Overfelt, *supra* note 220.

but customers are also abandoning products because of brand dilution through post-sale confusion.[264] Anna Corinna, co-founder of Foley + Corinna, reported that a bride-to-be returned over $1200 worth of bridesmaids dresses after seeing "the same dress" with "identical coloring, cut, and flower design" at Forever 21.[265] Just like Ms. Sakellis's scarves, the $300 originals were made of silk, while the $40 copies were made of cheaper polyester.[266] Ms. Corinna stated "[w]hen one of our designs gets knocked off, the dress is cheapened—customers won't touch it."[267]

These examples demonstrate the fallacy of two arguments against fashion design protection: first, that only famous designers would benefit from the design protection; and second, that designers with higher price tags are not injured by low-end knockoffs because they do not have the same clientele. Fashion design protection may actually benefit emerging designers more than their well-established counterparts. Currently, small designers often do not find relief from copyists because they lack the finances for the uphill legal battle they would face.[268] A law specifically prohibiting knockoffs might give their bark a little more bite, and by extension, discourage copyists from infringing in the first place.

Some opponents worry that design protection would increase the price of fashion, resulting in a product that the average American could not afford.[269] However, designers often license their runway lines to stores like Target or Wal-Mart at affordable prices.[270] If anything, piracy makes it more difficult for the original designers to market these more affordable versions. For example, Vera Wang licensed two versions of Chelsea Clinton's $20,000 wedding gown to David's Bridal for the retail prices of $599 and $1,050.[271] However, Allen Schwartz's knockoff will inevitably divert customers away from the officially licensed version.[272] This behavior may create a disincentive for famous designers to make diffusion lines in the future, thereby doing a disservice to customers who want the designer's name attached to their apparel.

---

264. *See id.*

265. *Id.*; *see also* Ferla, *supra* note 15, at G5.

266. *See* Overfelt, *supra* note 220.

267. *Id.*

268. *See* Binkley, *supra* note 251 ("Small designers face a particularly large burden; often, they lack deep pockets to chase down versions they find similar . . . ."); Kover, *supra* note 41 (stating that it was unusual for an emerging designer to receive a response or satisfaction from a copying complaint).

269. *See H.R. 5055 Hearing*, *supra* note 8, at 221 (statement of the American Free Trade Association).

270. *See id.* at 81 (statement of Susan Scafidi).

271. *See* Claire Howorth, *Chelsea's Wedding Dress Knockoff*, DAILY BEAST (Aug. 2, 2010, 5:59 PM), http://www.thedailybeast.com/blogs-and-stories2010-0802/chelseaclintons-wedding-dress-knockoff.

272. *See id.* While the licensed versions cost less than the $20,000 original, they are still "considerably higher than Schwartz's $100-plus range." *Id.*

Having knockoffs in the market can lead to the post-sale confusion of the public, and as a result, affect their purchasing decisions.[273] Post-sale confusion occurs when the public associates an infringing article with the original source.[274] When the market is flooded with knockoffs, "sales of the originals may decline because the public is fearful that what they are purchasing may not be an original."[275] Therefore, consumers who may have initially been willing to pay for the designer product may end up not doing so because they are afraid that it is not authentic.

Similarly, those who do regularly purchase high-end goods, or goods associated with a particular designer, may be harmed.[276] In *Hermes Int'l v. Lederer de Paris Fifth Ave., Inc.*, the court stated that "the purchaser of an original [design] is harmed by the widespread existence of knockoffs because the high value of originals, which derives in part from their scarcity, is lessened."[277] As a result, those consumers may be discouraged from buying in the future. This also leads to reverse confusion in which the public associates a knockoff product with the original, by extension harming the designer.[278]

Thus, this all comes back to the designer. For example, a ruffed sleeved canary-yellow harness worn by Willow Smith, daughter of actor Will Smith, to the 2010 American Music Awards initially received rave reviews.[279] However, it was quickly discovered that what she wore was not an original work—a terrible faux pas—and instead, was a knockoff of a design by Mildred Von Hildegard, the designer behind the label Mother of London.[280] When learning that a copy of her work had appeared on the red carpet, she stated that she was "disappointed and saddened that [it] will negatively impact [her] business. [Her] original design is now recognized in the mass media as someone else's work, and [it is] very hurtful on both a personal and professional levels [sic]."[281]

---

273. *See* Lois Sportswear v. Levi Strauss & Co., 799 F.2d 867, 872–873 (2d Cir. 1986); ROBERT P. MERGES, ET AL., INTELLECTUAL PROPERTY IN THE NEW TECHNOLOGICAL AGE 835–36 (5th ed. 2010).

274. *See Lois Sportswear*, 799 F.2d at 872–873; MERGES, ET AL., *supra* note 273, at 835–36.

275. *See* Hermes Int'l v. Lederer de Paris Fifth Ave., Inc., 219 F.3d 104, 108 (2d Cir. 2000).

276. *See id.*

277. *Id.*

278. *See* MERGES, ET AL., *supra* note 273, at 836–37.

279. *See Mother Of London Knocked Off at AMAs*, HAUTE MACABRE, (Nov. 22, 2010, 10:44 PM), http://hautemacabre.com/2010/11/mother-of-london-knocked-off-at-amas/comment-page-1/#comment-14374.

280. *See id.*

281. *Id.*

4. Similar Legislation in the European Union has Aided Fashion Designers in Protecting their Designs

To understand how fashion design protection would affect American designers, we can first look to similar laws existing in other countries and how those laws have influenced the industry. Legislation protecting fashion designs currently exists in Japan,[282] India,[283] Spain,[284] Italy,[285] France,[286] and the European Union.[287] The European Union's laws are most similar to those proposed in the United States because they provide protection for both registered designs, like the DPPA, and unregistered designs, like the IDPPPA.[288]

In 1998, the European Union adopted the European Directive[289] on the Legal Protection of Designs ("1998 Directive"), which created a "design right."[290] To receive protection, the design must be "new and [have]

---

282. *See* Isho Ho [Design Law], Law No. 125 of 1959, *translation available at* http://www.cas.go.jp/jp/seisaku/hourei/data/DACT.pdf.

283. *See* The Designs Act, No. 16 of 2000, INDIA CODE (2000), *translation available at* http://www.patentoffice.nic.in/ipr/design/design_act.PDF.

284. *See* Laura C. Marshall, *Catwalk Copycats: Why Congress Should Adopt a Modified Version of the Design Piracy Prohibition Act*, 14 J. INTELL. PROP. L. 305, 318 (2007) (citing Ben Smulders, *The European Community and Copyright § 4, in* INTERNATIONAL COPYRIGHT LAW & PRACTICE (Paul Edward Geller ed., 2006)) ("Designs that are registered with the Office for Harmonization in the Internal Market (OHIM) in Spain receive a five-year term of protection, renewable at five-year intervals for up to twenty-five years.").

285. *See* Emily S. Day, *Double-Edged Scissor: Legal Protection for Fashion Design*, 86 N.C. L. REV. 237, 267 (2007) (quoting Alberto Musso & Mario Fabiam, *Italy, in* INTERNATIONAL COPYRIGHT LAW AND PRACTICE § 2(4)(c)(i) (Paul Edward Gellar ed., 2006)) ("Finally, Italian copyright law extends protection to '[w]orks of industrial design displaying creative character and *per se* artistic value.'").

286. *See* Loi 94-361 du 10 mai 1994 [Law 94-361 of May 10, 1994, Copyright Act] JOURNAL OFFICIEL DE LA RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE], Aug. 2, 2003, p. 6863, art. L112-2, *unofficial version available at* http://natlawip.abra.info/european/france/frenchlegislation/prfr7.htm (Copyright Act includes "articles of fashion.").

287. *See* Council Regulation 6/2002, 2002 O.J. (L 3) 2 (EC), *available at* http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2002:003:0001:0024:EN:PDF, *amended by* Council Regulation 1891/2006, 2006 O.J. (L 386) 14 (EC), *available at* http://oami.europa.eu/en/design/pdf/l_38620061229en00140016.pdf. The changes made by the 2006 amendment are not relevant to this discussion.

288. *See* S. 3728, 111th Cong. § 2(f)(2) (2010); H.R. 2196, 111th Cong. § 2(f) (2009).

289. A European Union directive is a legislative act that requires the Member States to adapt their laws to achieve a specific result, but does not prescribe how to achieve that result. *See What are EU Directives?*, EUROPEAN COMMISSION (May 18, 2010), http://ec.europa.eu/community_law/introduction/what_directive_en.htm.

290. Council Directive 98/71, art. 1, 1998 O.J. (L 289) 28, 30 (EC), *available at* http://www.ipjur.com/data/981013DIR9871EC.pdf.

individual character."[291] Under the 1998 Directive, a design is considered new if "no identical design has been made available to the public" prior to the application's filing date and has individual character if the "overall impression" on the "informed user" differs from any preexisting design.[292] The 1998 Directive protects the design's overall appearance as well as any part arising from "the lines, contours, colours, shape, texture and/or materials of the product itself and/or its ornamentation."[293] If an article receives protection under the 1998 Directive, it is protected for up to twenty-five years from its filing date.[294]

In 2002, the European Union Council enacted a regulation providing additional rights for designs with the European Regulation[295] on Community Designs ("2002 Regulation").[296] The 2002 Regulation offers protection similar to that of the 1998 Directive, except that it provides a dichotomy of protection for registered and unregistered designs.[297] Moreover, the 2002 Regulation clarified that, because its purpose was to prevent copying, its protections would not prohibit works that were born of independent creation.[298] Finally, registered designs are still protected for twenty-five years under the 2002 Regulation, but unregistered designs are limited to a term of three years.[299]

Since the 2002 Regulation was enacted, the number of apparel design registrations has risen from just 3,156 registrations in 2003[300] to 8,192 in 2009,[301] for a total of 42,888 registered apparel designs.[302] David Musker, an intellectual property attorney and leading expert on designs,[303] believes

---

291. *See id.* art. 3, at 30.

292. *Id.* art. 4-5, at 30.

293. *Id.* art. 1, at 30.

294. *See id.* art. 10, at 31.

295. A European Union regulation is a legislative act that becomes immediately binding as law in all Member States. *What are EU Regulations?*, EUROPEAN COMMISSION (May 18, 2010), http://ec.europa.eu/community_law/introduction/what_regulation_en.htm.

296. *See* Council Regulation 6/2002, *supra* note 287.

297. *See id.* at 2.

298. *See id.* at 3.

299. *See id.* art. 11-12, at 5.

300. *SSC007 - Statistics of Community Designs 2003*, 8 (2004), *available at* http://oami.europa.eu/pdf/office/Stats%20RCD%202003.pdf; *see* WIPO, *List of Goods in Class Order*, 4, *available at* http://www.wipo.int/classifications/nivilo/pdf/eng/locarno/loc2eng.pdf (classifying goods for regulation purposes).

301. *SSC007 - Statistics of Community Designs 2009*, 13 (2010), *available at* http://oami.europa.eu/ows/rw/resource/documents/OHIM/statistics/ssc007-statistics_of_community_designs_2009.pdf.

302. *SSC007 - Statistics of Community Designs 2010*, 11 (2010), *available at http://oami.europa.eu/ows/rw/resource/documents/OHIM/statistics/ssc007-statistics_of_community_designs_2010.pdf.*

303. *Rankings*, JENKINS | TRADEMARK & PAT. ATT'YS, http://www.jenkins.eu/about-us/rankings.asp (last visited Jan. 1, 2011).

that the European Union's legislation "has enabled design [registration] to be used as an anti-piracy weapon."[304] Specifically, Mr. Musker notes that the laws are effective for "quick copying industries" in which designs are often replicated and distributed before the originals are in stores.[305]

Two designers who have found relief through the 2002 Regulation are Jimmy Choo and Karen Millen.[306] Jimmy Choo sued Towerstone Ltd. for copying one of the large bucket bags he registered as a community design.[307] Jimmy Choo prevailed when the judge found that an informed user, not just a "woman in the street," would find the bags "exactly the same."[308] The judge made this determination despite differences in the fabric's texture, the clasp, and the number of eyelets through which a belt of fabric ran at the top of the bag.[309]

Similarly, Karen Millen found relief for three unregistered designs consisting of "a black knit top, a blue shirt and a brown shirt."[310] Ms. Millen sued Dunnes Stores for selling copies of her designs.[311] Dunnes Stores admitted copying the designs but claimed they were not liable for infringement because Ms. Millen's designs lacked individual character, making them ineligible for protection.[312] The court found that the items had individual character, despite considering photos of similar shirts produced by Dolce & Gabana and Paul Smith submitted by the defendant.[313] Because the Karen Millen tops had individual character, the court ordered Dunnes Stores to stop selling the replicas.[314]

Under the current United States law and previous European law, these outcomes would likely have been impossible. Gary Assim, an English intellectual property attorney, notes that in the European Union, "[t]raditionally, [one] had to prove that an item of clothing or footwear was identical, which is close to impossible, but the new [European Union]

---

304. OHIM, *Annual Report 2009*, 8 (2009), *available at* http://oami.europa.eu/en/Annual_Report_2009/content/pdfs/AR2009_EN.pdf.

305. *Id.* Mr. Musker further stated: "I think the Community design system, which links well with the EU's customs measures, has enabled designs to be used as an anti-piracy weapon." *Id.*

306. *See* J Choo (Jersey) Ltd. v. Towerstone Ltd., [2008] EWHC 346 (Ch) (H. Ct.), *available at* http://oami.europa.eu/pdf/design/cdcourts/Handbags.pdf; Karen Millen Ltd. v. Dunnes Stores, [2007] IEHC 449 (H. Ct.), *available at* http://oami.europa.eu/pdf/design/cdcourts/Ireland.pdf.

307. *See J Choo (Jersey) Ltd.*, [2008] EWHC 346 (Ch) at 1–2.

308. *Id.* at 2, 4.

309. *See id.* at 3–4.

310. *Karen Millen Ltd.*, [2007] IEHC 449 at 2.

311. *See Karen Millen Ltd.*, [2007] IEHC 449.

312. *See id.* at 15.

313. *See id.* at 22–23.

314. *See id.* at 24.

legislation allows you to take an aspect of design, such as a quirky buckle or a heel, and to show that it has been copied."[315]

5. Similar Legislation in the European Union Does Not Indicate the Likelihood of Decreased Creativity or an Impeded Design Cycle

Some scholars argue that copying drives the fashion industry because it forces designers to respond with new ideas to set themselves apart.[316] As a corollary, they assert that design protection would slow the fashion industry.[317] However, the fashion industries in countries with design protection also serve as evidence that enacting similar legislation in the United States will not hamper innovation.[318] Professor Scafidi suggests that, because other countries have woven fashion design into their intellectual property laws, they have "developed more mature and influential design industries."[319] In fact, France, the global fashion leader, has some of the most substantial design protection in the world.[320] Stricter design protection in the United States may actually increase innovation.[321] Not only will designers be able to create freely with the promise of protection for their original works, but pure copyists will have to add original elements to their own designs to avoid infringing upon the protected works of other designers.[322] In other words, the industry will grow through design innovation because designers will be prohibited from selling identical products.[323]

6. Similar Legislation in the European Union Does Not Indicate the Likelihood of Increased Litigation

Christopher Sprigman, a professor at the University of Virginia School of Law and constant opponent of copyright expansion of all types,[324] argues that the enactment of laws protecting fashion design "might turn the

---

315. *New Look Withdraws 1,000 Shoes to Settle Copying Case*, LONDON TIMES, Sept. 13, 2006, *available at* http://business.timesonline.co.uk/tol/business/industry_sectors/retailing/article637109.ece.

316. *See* Raustiala & Sprigman, *supra* note 238, at 1722.

317. *See id.*

318. *See H.R. 5055 Hearing*, *supra* note 8, at 83–84 (statement of Susan Scafidi); Scafidi, *supra* note 49, at 117.

319. *H.R. 5055 Hearing*, *supra* note 8, at 83 (statement of Susan Scafidi).

320. *See id.* at 84 (statement of Susan Scafidi); Day, *supra* note 285, at 266.

321. *See* STOP FASHION PIRACY, *supra* note 9.

322. *See id.*

323. *See id.*

324. In addition to fashion design protection, Professor Sprigman has opposed "the Copyright Renewal Act of 1992, the Sonny Bono Copyright Term Extension Act, the Copyright Act of 1976, [and] the Berne Convention Implementation Act." *H.R. 5055 Hearing*, *supra* note 8, at 6 (statement of Rep. Bob Goodlatte).

industry's attention away from innovation and toward litigation."[325] While Professor Sprigman concedes that European Union fashion designers have not responded litigiously since their designs were granted protection, he credits this to the differences in the European and United States legal systems.[326]

The limited litigation that has occurred in the European Union since its enactment of the 2002 Regulation has generally focused on "copycat fashion" rather than designer-on-designer brawls.[327] This is similar to the type of litigation already seen in the United States against companies, like Forever 21, which thrive on copying designs.[328] Because the DPPA and the IDPPPA target the type of behavior that is already the subject of litigation in the United States, they may instead decrease lawsuits by banning such behavior. In fact, Alain Coblence, an attorney who helped develop S. 1957, stated that similar legislation in other countries actually decreased lawsuits.[329] He stated that, "'[w]hen you look at what is happening in countries that have laws against copying, . . . there are few lawsuits because they know it is illegal and they don[']t do it.'"[330] Explicit design protection would make it more likely that a designer would prevail, thus reducing the expected value of copying. This is counter to the current system in which the chance of prevailing is less certain and the costs of copying are correspondingly lower.

Professor Sprigman also argues that the laws in the European Union have been ineffectual because designers have not utilized the registration system and piracy continues to occur in the European fashion industry.[331] However, the law is intended to be used as a shield and not as a sword, and those fashion designers who have elected to pursue their rights have indeed found relief.[332]

B. What Do The DPPA and IDPPPA Mean?

This section will analyze and compare key elements of the DPPA and the IDPPPA, how the bills address the concerns of opponents and

---

325. *Id.* at 88 (statement of Christopher Sprigman).

326. *See id.* ("Unlike most countries in Europe, which have relatively weak civil litigation systems, we Americans are, for better or worse, accustomed to resolving disputes through the courts.").

327. *See* Karen Fong & Tom Grek, *IP Special Report: Crimes of Fashion*, LAWYER (Jan. 19, 2009), at 21, *available at* http://www.thelawyer.com/ip-special-report-crimes-of-fashion/136319.article.

328. *See* Odell, *supra* note 13.

329. *See Fashion Fission over Anti-Knockoff Bill*, L.A. BUS. J. (Sept. 10, 2007), *available at* http://www.buchalter.com/bt/index.php?option=com_content&task=view&id=183&Itemid=57.

330. *Id.*

331. *See H.R. 5055 Hearing*, *supra* note 8, at 88 (statement of Christopher Sprigman).

332. *See supra* Part III.A.4.

proponents, the bills' potential effects on the fashion industry, and will propose solutions to problematic language contained in the bills. The DPPA and IDPPPA would amend Chapter 13 of the Copyright Act to provide three years of limited protection to original fashion designs.[333] The DPPA would exclude protection for designs that constituted "trends" and would require registration before they could be protected.[334] The IDPPPA would not have a registration requirement, but would raise the bar on what would be considered "original" and require a pleading before a lawsuit could go forward.[335] Senator Schumer stated that the IDPPPA struck "the right balance between providing narrow protection to truly unique and original fashion designs while deterring and preventing frivolous lawsuits."[336]

1. Limitations on Designs Protected

a. Similarities Between the DPPA and IDPPPA

Both bills would make the necessary addition of "an article of apparel" under the definition of useful articles in Chapter 13 the Copyright Act.[337] The bills would amend § 1301(a)(3) of the Copyright Act to include fashion as a protected category under the *sui generis* design protection located in Chapter 13 of the Copyright Act,[338] a section of the Copyright Act currently limited to protecting boat hull design.[339] In determining whether a design could obtain protection, each fashion design would be considered as a whole and would only include the original elements and their placement "in the overall appearance of the article of apparel."[340] This is an important component to the proposed legislation because it explicitly excludes articles that are merely replicas of past designs and those that are wardrobe classics, such as plain black cardigans, which do not have an original design or detail.[341] Because this revision already exists within the statute, it would

333. *See* S. 3728, 111th Cong. (2010); H.R 2196, 111th Cong. (2009).

334. *See* H.R 2196, 111th Cong. (2009).

335. *See* S. 3728, 111th Cong. (2010).

336. Executive Business Meeting, *supra* note 212, at 52:00. Senator Orrin Hatch, a co-sponsor of the IDPPPA, commented that it was an "important bill" that would "provide very limited protection to fashion designs that are so original that they approach art, rather than utilitarian fashion." Executive Business Meeting, *supra* note 212, at 52:50.

337. *See* S. 3728, 111th Cong. § (2)(a)(2) (2010); H.R. 2196, 111th Cong. § 2(a)(2) (2009); *see also* 17 U.S.C. § 1301(b)(2) (2006).

338. *See* S. 3728, 111th Cong. § (2)(a)(1) (2010); H.R 2196, 111th Cong. § (2)(a)(1) (2009); *see also* 17 U.S.C. § 1301(a)(3) (2006).

339. 17 U.S.C. § 1301 (2006).

340. S. 3728, 111th Cong. § 2(a)(2) (2010); H.R. 2196, 111th Cong. § 2(a)(2) (2009); *see also H.R. 5055 Hearing*, *supra* note 8, at 197 (statement of the United States Copyright Office).

341. *See* S. 3728, 111th Cong. § (2)(a)(2) (2010); H.R. 2196, 111th Cong. § 2(a)(2) (2009).

make it clear to the courts that a fashion design's ornamental and design elements should be considered separately from its utilitarian function and that designs merely dictated by utilitarian function would not be protected.[342]

The revisions are fairly comprehensive, defining apparel as: "(A) an article of men's, women's, or children's clothing, including undergarments, outerwear, gloves, footwear, and headgear; (B) handbags, purses, wallets, duffel bags, suitcases, tote bags, and belts; and (C) eyeglass frames."[343] This list expands upon previous bills, which only listed "handbags, purses, and tote bags" as the types of carrying cases that would be protected.[344] The present version's enumeration of these categories may be problematic by allowing infringers to claim that items not specifically mentioned, such as briefcases or scarves, were not intended for protection. To prevent this potential loophole, supporters should clarify the legislative intent by revising the definition of "apparel" to either to limit it expressly to the listed articles or incorporate language that would make the list nonexclusive and include like articles.[345]

b. Additions Proposed by the IDPPPA

The IDPPPA adds to the originality requirement, requiring that the elements of the design "(i) are the result of a designer's own creative endeavor; and (ii) provide a unique, distinguishable, non-trivial and non-utilitarian variation over prior designs for similar types of articles."[346] This language has led to fears that creating a design that is entitled to protection would be impossible.[347] However, it may just be a clarification of legislative intent or an attempt to appease opponents by adding language to provide a narrower, more focused definition of "original."[348] Chapter 13 of

---

342. *See* Joseph E. McNamara, *Modifying the Design Piracy Prohibition Act to Offer "Opt-Out" Protection for Fashion Designs*, 56 J. COPYRIGHT SOC'Y U.S.A. 505, 518 (2009).

343. S. 3728, 111th Cong. § 2(a)(2) (2010); H.R. 2196, 111th Cong. § 2(a)(2) (2009).

344. *Compare* S. 1957, 110th Cong. § 2(a)(2) (2007), *and* H.R. 2033, 110th Cong. § 2(a)(2) (2007), *and* H.R. 5055, 109th Cong. § 1(a)(2) (2006), *with* S. 3728, 111th Cong. § 2(a)(2) (2010), *and* H.R. 2196, 111th Cong. § 2(a)(2) (2009).

345. Section 102 of the Copyright Act, which defines the subject matter to which protection is available, includes language that extends protection to "works of authorship fixed in any tangible medium of expression, now known or later developed." 17 U.S.C. § 102 (2006). The DPPA or the IDPPPA could incorporate the "now known or later developed" language to close this loophole.

346. S. 3728, 111th Cong. § 2(a)(2) (2010).

347. Staci Riordan, *Breaking News: New Design Piracy Bill Introduced into Sentate* [sic], FASHION L. BLOG (Aug. 6, 2010, 4:22 PM), http://fashionlaw.foxrothschild.com/2010/08/articles/design-piracy-prohibition-act/breaking-news-new-design-piracy-bill-introduced-into-sentate.

348. Professor Scafidi, who helped develop the IDPPPA, indicated that the bill's conception of originality was the result of extensive negotiations with former opponents. *See*

the Copyright Act already requires the designs to be original and non-utilitarian in nature.[349] The "non-trivial" requirement seems to rise above the normal minimum requirements for creativity,[350] and this additional language could create problems for judges and juries who must assume the role of "ad hoc fashion experts" and use their own interpretations of what "distinguishable" and "non-trivial" mean in terms of apparel.[351] This analysis seems more akin to the nonobvious requirement found in patent law than the lower modicum of creativity standard found in copyright law.[352] The IDPPPA's attempt to further define "fashion design" only raises more questions as to what it would actually protect. Because much of the language is redundant, it may be better to omit the additional parameters under the fashion design definition.

The IDPPPA also adds a rule of construction, which dictates that "differences or variations which are considered non-trivial for the purposes of establishing that a design is subject to protection . . . shall be considered non-trivial for the purposes of establishing that a defendant's design is not substantially identical" under an infringement analysis.[353] This provision essentially places plaintiffs and defendants on the same footing by ensuring that the design's elements are interpreted consistently for both parties.

c. Proposed Compromise

Some opponents of the DPPA fear that its language would allow a designer to monopolize "something as common as a hoodie."[354] However, a simple reading of the relevant language of the current Chapter 13 of the Copyright Act in conjunction with the DPPA reveals such fears to be baseless, as together they plainly state that only original designs would be afforded protection.[355] Because the DPPA would not protect non-original pieces, IDPPPA proponents may have developed the additional originality

---

Susan Scafidi, *IDPPPA: Introducing the Innovative Design Protection and Piracy Prevention Act, a.k.a. Fashion Copyright*, COUNTERFEIT CHIC (Aug. 6, 2010) [hereinafter Scafidi, *IDPPPA*], http://counterfeitchic.com/2010/08/introducing-the-innovative-design-protection-and-piracy-prevention-act.html; *see also H.R. 5055 Hearing*, *supra* note 8, at 16 (statement of David Wolfe, creative director of The Doneger Group) (stating that it is difficult to define originality).

349. *See* 17 U.S.C. § 1302(1), (4) (2006).

350. *See* Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 346 (1991) (requiring a "modicum of creativity").

351. Brown, *supra* note 23.

352. *See supra* Parts II.B.2, II.B.4.

353. S. 3728, 111th Cong. § 2(a)(3) (2010).

354. Kathleen Fasanella, *IDPPPA: Yet Another Fashion Design Copyright Law*, FASHION INCUBATOR (Aug. 6, 2010, 7:09 AM), http://www.fashion-incubator.com/archive/idpppa-yet-another-fashion-design-copyright-law.

355. *See* 17 U.S.C. § 1302(1) (2006) (designs that are not original would not be subject to protection).

requirements out of compromise instead of necessity.[356] The problem with the IDPPPA's language in § 2(a)(2) is that it would cover particularly novel pieces but may not cover the glamorous one-of-a-kind red carpet gowns featured at award shows.[357] For example, while Jean Charles de Castelbajac's Kermit the Frog jacket[358] might obtain protection (minus the obvious trademark issues), a Valentino gown, such as the one Julia Roberts wore to the 74th Annual Academy Awards,[359] might not. With its current language, the IDPPPA would provide little protection from fashion pirates like Allen Schwartz and Forever 21 and would therefore be virtually useless in fighting piracy. Congress should either define key IDPPPA terms, such as "unique," "distinguishable," "non-trivial," and "non-utilitarian," or adopt the more succinct clause found in the DPPA.[360]

2. Revisions, Adaptations, and Rearrangements

The differences in language between the DPPA and the IDPPPA in this section are minor and will therefore be discussed together. Designs that use subject matter excluded from protection by § 1302, such as those that are commonplace or not original, are protected if they have "a substantial revision, adaptation, or rearrangement of such subject matter."[361] The existing revision allowance in § 1303 of the Copyright Act permits designers to build on past designs by adding their own original, creative aspects.[362] Because almost all fashion builds on that which came before it, this provision is very important.

The bills would amend § 1303 of the Copyright Act to provide that the addition or absence of colors or graphics on the material would not be a factor in determining either the protection or infringement of fashion

---

356. *See* Sheppard Mullin, *The Innovative Design Protection and Privacy Prevention Act: Will Design Protection Be in Vogue in Congress?*, FASHION APPAREL L. BLOG (Aug. 23, 2010, 7:07 AM), http://www.fashionapparellawblog.com/2010/08/articles/enforcement-of-fashion-laws/the-innovative-design-protection-and-privacy-prevention-act-will-design-protection-be-in-vogue-in-congress.

357. *See* S. 3728, 111th Cong. § 2(a)(2) (2010); Cathy Horyn, *Schumer Bill Seeks to Protect Fashion Design*, N.Y. TIMES (Aug. 5, 2010, 10:43 PM), http://runway.blogs.nytimes.com/2010/08/05/schumer-bill-seeks-to-protect-fashion-design.

358. *See Kermit Coat? Muppets Invade French Fashion*, MSNBC (Mar. 10, 2009, 3:51 PM), http://today.msnbc.msn.com/id/29618702.

359. Anita Singh, *Valentino Pays Tribute to Julia Roberts at Venice Film Festival*, TELEGRAPH (Aug. 28, 2008, 3:13 PM),http://www.telegraph.co.uk/news/newstopics/celebritynews/2638802/Valentino-pays-tribute-to-Julia-Roberts-at-Venice-Film-Festival.html.

360. *See* S. 3728, 111th Cong. § 2(a)(2) (2010); H.R. 2196, 111th Cong. § 2(a)(2) (2009).

361. 17 U.S.C. § 1303 (2006); *see* § 1302.

362. *See id.* § 1303.

designs.[363] The exclusion of color and graphics from an infringement analysis could have both positive and negative effects. On the one hand, it could limit what is considered an infringing article and "reduce the complications in determining whether a design infringes."[364] On the other hand, in some cases, like the Louis Vuitton Murakami bags, it could leave little to the infringement analysis, and perhaps even reduce the designs to their structures or even their most utilitarian forms. A possible rationale for this language may be that, because copyright law already protects prints on fabric, making them a major factor in the similarity or infringement analysis is not as important.[365]

It seems that Congress intends this language to mean that a designer cannot claim protection for an existing design solely by changing the colors used. Conversely, a copyist cannot escape infringement by making a protected design in a different color. Alternatively, two otherwise different pieces of apparel would not infringe upon one another simply because they happen to be the same color or use a similar fabric. For example, a cerulean military-style jacket would not infringe upon a cerulean ball gown.

However, the language could be read very broadly, leading to a situation in which color and graphics are never considered as parts of the design. This language should be clarified or changed to allow the use and arrangement of color and graphics in both the protection and infringement analyses, but it should stipulate that the use of color or graphics in and of themselves are not enough to establish protection or infringement on its face.

3. Publication of Designs

Both bills use the design's date of publication to start its term of protection.[366] A design is made public when a "useful article embodying the design is . . . publicly exhibited, . . . distributed, or offered for sale or sold to the public by the owner of the design or with the owner's consent."[367]

---

363. *See* S. 3728, 111th Cong. § 2(c) (2010); H.R. 2196, 111th Cong. § 2(c) (2009). The DPPA provides that these factors would not be used when determining the "originality" of the design, while the IDPPPA uses the word "protection." Additionally, the DPPA provides that the factors would not be considered in "the similarity or absence of similarity of fashion designs" when determining infringement. *See id.* Though the language is different, the effect of the provisions is the same.

364. Kimberly Ann Barton, Note, *Back to the Beginning: A Revival of a 1913 Argument for Intellectual Property Protection for Fashion Design*, 35 J. CORP. L. 425, 438 (2009).

365. *See* Peter Pan Fabrics, Inc. v. Brenda Fabrics, Inc., 169 F. Supp. 142, 143 (S.D.N.Y. 1959).

366. *See* S. 3728, 111th Cong. § 2(b)(3) (2010); H.R. 2196, 111th Cong. § 2(b)(3) (2009).

367. 17 U.S.C. § 1310(b) (2006); *see also* S. 3728, 111th Cong. § 2(b)(3) (2010); H.R. 2196, 111th Cong. § 2(b)(3) (2009). The DPPA would include designs offered for individual or public sale. *See* H.R. 2196, 111th Cong. § 2(f)(2) (2009).

This last requirement is important in that the time period starts only with the publication by the owner or designer and not with publication by any party. This requirement will prevent others' unauthorized acts from starting the clock and thereby preventing the owner from fully benefiting from the statutory time frame. For example, if an overzealous Chanel intern posts a photograph of a Karl Lagerfeld[368] design to a social networking website long before its scheduled runway debut, that intern would not have "published" it under the DPPA or the IDPPPA.

Section 1310 also utilizes the phrase "existing useful article embodying the design,"[369] which would exclude the use of sketches and other renderings as a form of publication.[370] The development of a design is a lengthy process, often taking years from when the initial idea is put down on paper until the finished product is ready to hit the store or runway.[371] The requirement that the design be embodied in the useful article prevents the clock from running while the articles are still being made, which otherwise could leave little protection time once the designs are ready for display or sale.

Because the publication date affects the term of protection, this could change designers' preseason marketing methods.[372] For example, Lubov Azria, creative director of BCBGMAXAZRIAGROUP, often posts images of photo shoots and upcoming merchandise as a way to pique consumer interest before a runway debut.[373] While this method announces coming trends and products, it would also constitute a publication under the DPPA and the IDPPPA if the image is of a fully formed article of apparel.[374] One potential result may be that designers will delay advertising to ensure that they receive the maximum benefit of the term of protection. On the other hand, the revenue produced by the additional promotion may be more

---

368. Karl Lagerfeld is Chanel's artistic director. *See* John Colapinto, *Where Karl Lagerfeld Lives*, NEW YORKER, Mar. 19, 2007, at 112, *available at* http://www.newyorker.com/reporting/2007/03/19/070319fa_fact_colapinto.

369. 17 U.S.C. § 1310(b) (2006).

370. Because a drawing's purpose is to "portray the appearance" of the design, it is not a useful article. *See id.* § 101.

371. BUREAU OF LABOR STATISTICS, U.S. DEP'T OF LABOR, *supra* note 222, at 1. ("The design process from initial design concept to final production takes between 18 and 24 months.").

372. *See* Fasanella, *supra* note 354 (discussing a different—and wrongly interpreted—section proposed by the IDPPPA).

373. *See* Lubov Azria, *Summer Fantasy: The Summer 2010 Photo Shoots Keep Us Warm on the Coldest Day of the Year So Far*, BCBGMAXAZRAIGROUP (Feb. 1, 2010, 12:00 PM), http://bcbgmaxazriagroup.com/fall2009/index.php?p=cGFnZT1mdWxsX2Jsb2cmaWQ9MzM0&lnt=cmVhZF9tb3Jl&section=fashionfiles&cat=4 (posting images of upcoming summer designs).

374. *See* 17 U.S.C. § 1310(b) (2006) ("A design is made public when an existing useful article embodying the design is anywhere publicly exhibited . . . .").

appealing to designers in the long run, minimally affecting their marketing strategies—if it affects them at all.

4. Term of Protection

Under the proposed bills, a fashion design would receive protection for a nonrenewable three-year term.[375] While this is much shorter than the ten-year period currently provided to vessel hull designs under Chapter 13 of the Copyright Act,[376] the term strikes a good balance between protecting designers from the fast-fashion knockoff artists and accommodating the fashion industry's naturally rapid turnover.[377] Because the piracy process has become faster as technology has progressed, the three-year term essentially levels the playing field for designers, in effect returning fashion to the traditional "quick but not instant" copying regime that has served the industry well for so long. In other words, it is a good fit for fashion. In fact, the Copyright Office "applaud[ed]" the "modest term" of protection proposed by the bill's proponents, noting that it was "calibrated to address the period of time during which fashion designs are most at risk of being infringed."[378] After the short three-year period, the designs would become public domain and could be used freely by other designers. Because fashion has a limited shelf life, this would likely not adversely affect the original designer finances.

5. Registration

a. The DPPA

To receive copyright protection under the DPPA, a fashion design must be registered within six months of the date it is made public.[379] Because of the ever-changing nature of the fashion industry, this term is much shorter than the two-year grace period allowed for designs of vessel hulls under current Chapter 13.[380] This short period would prevent designers from sitting on their rights and is consistent with the fast-paced nature of the design world.

As part of the registration, the designer or owner must provide a "brief description of the design," for the database.[381] This information would be

---

375. *See* S. 3728, 111th Cong. § 2(d) (2010); H.R. 2196, 111th Cong. § 2(d) (2009).

376. 17 U.S.C. § 1305(a) (2006).

377. *H.R. 5055 Hearing*, *supra* note 8, at 210 (statement of the United States Copyright Office).

378. *Id.* (statement of the United States Copyright Office) (referring to H.R. 5055 but proposing the same term of protection as that in the current H.R. 2196).

379. H.R. 2196, 111th Cong. § 2(f)(1) (2009).

380. 17 U.S.C. § 1302(5) (2006).

381. *See* H.R. 2196, 111th Cong. § 2(f)(3), (j) (2009).

used solely to aid others in finding like designs and would not "limit the protection granted to the design."[382] The registration fees are not specifically set forth, but are expected to be approximately $100 per design.[383]

The DPPA would also create an electronic database, searchable by categories of apparel and accessories, for fashion designs.[384] The database would be free to the public, making it accessible to designers of all income levels.[385] The database would require "visual representation[s]" of the fashion designs and would list the items' registration status.[386]

b. The IDPPPA

The IDPPPA has no registration requirement for fashion designs.[387] Under the IDPPPA, if a fashion design is made public before either the bill's enactment "or more than 3 years before the date upon which protection of the design is asserted," then the article would not receive protection.[388] This not only eliminates the time and money that it would take to register articles, but also limits any potential attorneys' fees associated with the application process. The absence of a registration requirement would be particularly beneficial to start-up designers who do not have disposable income. Moreover, the 2002 Regulation in the European Union has demonstrated that designers can successfully find relief in a non-registration-based system.[389]

c. Proposed Compromise

The benefit of the electronic database proposed in the DPPA is that it provides notice that a design is protected.[390] This informs others that copying the design is illegal.[391] Although not having a registration system would save money for emerging designers, it might have the unintended result of complicating fashion design protection in general by making it

382. *Id.* § 2(f)(3).

383. *See* 17 U.S.C. § 1310(j) (2006); Overfelt, *supra* note 220.

384. *See* H.R. 2196, 111th Cong. § 2(j) (2009).

385. *See id.*

386. *Id.*

387. *See* S. 3728, 111th Cong. § 2(f)(2) (2010).

388. *Id.* § 2(b)(3).

389. *See* Council Regulation 6/2002, *supra* note 287, at 2; Karen Millen Ltd. v. Dunnes Stores, [2008] IEHC 449 (H. Ct.),
available at http://oami.europa.eu/pdf/design/cdcourts/Ireland.pdf; *supra* Part 4

390. *See* H.R. 2196, 111th Cong. § 2(j) (2009).

391. *See id.*; Steven Zuccarelli, *Fashion IP Revisited: The Innovative Design Protection and Piracy Prevention Act*, OSGOODE (Aug. 24, 2010), http://www.iposgoode.ca/2010/08/fashion-ip-revisited-the-innovative-design-protection-and-piracy-prevention-act.

more difficult to prove access and copying.[392] Therefore, Congress should incorporate the searchable electronic database while making registration accessible to new designers by setting the registration costs at the lowest possible rate that would cover the costs of registration and database maintenance.

6. Infringement

a. Acts with Knowledge

Under the current Chapter 13 of the Copyright Act, it is only an infringing act to knowingly make, sell, or distribute a protected design.[393] Though the DPPA and the IDPPPA propose different amendments to the "acts with knowledge" provision, they are carefully crafted and well-suited for their respective situations.[394]

i. The DPPA

The language of the DPPA would include individuals or companies who made duplicate designs and had "reasonable grounds to know that protection for the design [was] claimed."[395] This is targeted toward those who intentionally copy others' works, and not an average independent designer. The restriction allows the owner of infringed designs to claim that the infringer should have known that the work was protected. Arguably, because the registration system provided for in § 2(j) of the DPPA would be in place, designers would be on constructive notice that a registered design is protected. This would be akin to trademark law, which utilizes a searchable database of all registered marks to provide notice of their protection.[396] All one would need to do is search the database to see exactly which designs within a specific category have been registered.[397]

ii. The IDPPPA

The IDPPPA would include on the list of acts constituting infringement those in which the accused infringer knowingly advertised and made an offer for sale.[398] Additionally, it would amend Chapter 13 to specify that knowledge could be "either actual or reasonably inferred from the totality

392. *See* Zuccarelli, *supra* note 391.
393. *See* 17 U.S.C. § 1309(c) (2006).
394. *See* S. 3728, 111th Cong. § 2(e) (2010); H.R. 2196, 111th Cong. § 2(e) (2009).
395. H.R. 2196, 111th Cong. § 2(e) (2009); *see also* 17 U.S.C. § 1309(c) (2006).
396. *See* 15 U.S.C. §1072 (2006).
397. *See* H.R. 2196, 111th Cong. § 2(j) (2009).
398. *See* S. 3728, 111th Cong. § 2(e) (2010).

of the circumstances."[399] This standard requires a plaintiff to show that the defendant knew, not merely should have known, that he was doing something improper. However, because knowledge can be inferred through circumstantial evidence it may aid a plaintiff in convincing a judge or jury of infringement. To draw from copyright, a plaintiff may be able to demonstrate knowledge if the design had "been widely disseminated to the public" via advertising or media coverage.[400] Alternatively, knowledge could be inferred if the original and infringing articles were so similar that they could not have been created independently.[401] Because evidence of direct copying can be difficult to find, the use of circumstantial evidence is key in allowing an infringed designer to obtain relief.

b. Infringing Article Defined

Both bills would expand the definition of an infringing article to include copies made from images of the original, instead of being limited to copies made from the original itself.[402] Although this expansion may just be a clarification of the current statutory language (as the Copyright Office has suggested would be helpful), it is extremely important in the fashion industry since many duplicates are made from images of models on the runway or celebrities on the red carpet wearing the garment rather than from the article itself.[403] Both bills include an independent-creation exception that is on par with current copyright protection.[404] To exclude this exception would run counter to the intent behind the proposed legislation—as well as copyright law in general.[405] The purpose of these amendments is to prohibit profiteering based on others' work, not to inhibit those who independently create.[406]

i. The DPPA

Under the revisions of the DPPA, a design would not infringe upon a protected fashion design if it is "not closely and substantially similar in

399. *Id.*

400. Selle v. Gibb, 741 F.2d 896, 901 (7th Cir. 1984) (citing Abkco Music, Inc. v. Harrisongs Music, Ltd., 722 F.2d 988, 988 (2d Cir.1983)).

401. *See id.* at 900.

402. *Compare* S. 3728, 111th Cong. § 2(e) (2010) (defining an infringing article as a design copied from an original or "from an image thereof"), *and* H.R. 2196, 111th Cong. § 2(e)(2) (2009) (finding infringement from copies of design or "images thereof"), *with* 17 U.S.C. § 1309(e) (2006) (stating infringement for articles "copied from a design" but not mentioning images).

403. *See H.R. 5055 Hearing*, *supra* note 8, at 212–13 (statement of the United States Copyright Office).

404. *See* 20 AM. JUR. 3D *Proof of Facts* § 2 (1993).

405. *See* S. 3728, 111th Cong. (2010); H.R. 2196, 111th Cong. (2009).

406. *See* S. 3728, 111th Cong. (2010); H.R. 2196, 111th Cong. (2009).

overall visual appearance."[407] This is a higher standard than the "substantially similar" requirement under the current version of Chapter 13 of the Copyright Act[408] but not as high as the line-by-line standard proposed by some critics.[409] A problem with the "closely and substantially similar" language is that it creates an entirely new legal standard without any history to suggest its likely impact. While it is evident that the DPPA's proponents are trying to make compromises with its critics,[410] courts would likely prefer the consistency and continuity a more a familiar test would afford them.

This section would also make an exception for designs that "merely reflect[] a trend."[411] This applies to "newly popular concept[s], idea[s], or principle[s]" that create an "immediate amplified demand for articles" that embody those styles.[412] Earlier versions did not contain this limitation and caused concerns that they would "prohibit the ability of designers and retailers to replicate current trends and styles."[413] This language seems to be a compromise with the bill's opponents.[414] However, it could inadvertently block trendsetters from gaining protection by barring protection for the designs that launched the newly popular fashion. If a design is truly unique, it should be protected, regardless of its participation in a trend. Because this provision seems redundant and serves no legitimate purpose, it would be better to exclude it. Maintaining it would merely create extra language for courts to interpret, potentially creating more hurdles for designers to overcome.

ii. The IDPPPA

The IDPPPA proposes the even higher standard of "substantially identical," which is limited to items that are "so similar in appearance as to be likely to be mistaken for the protected design, and contain[] only those differences in construction or design which are merely trivial."[415] This

---

407. H.R. 2196, 111th Cong. § 2(e)(2) (2009).

408. *See* 17 U.S.C. § 1309(e) (2006).

409. *See H.R. 5055 Hearing*, *supra* note 8, at 7 (statement of Christopher Sprigman) ("I think it would be better if the bill were clearly limited only to those garments that are point-by-point copies of existing garments, but I don't think that is necessary either, even though it would clearly be better than what we have now.").

410. Scafidi, *IDPPPA*, *supra* note 348.

411. H.R. 2196, 111th Cong. § 2(e)(2) (2009).

412. *Id.* § 2(a)(2).

413. *H.R. 5055 Hearing*, *supra* note 8, at 2 (statement of Rep. Jim Sensenbrenner); *see* S. 1957, 110th Cong. § 2(a)(2) (2007); H.R. 2033, 110th Cong. § 2(a)(2) (2007); H.R. 5055, 109th Cong. § 1(a)(2)(B) (2006).

414. Representative Goodlatte asked DPPA opponent David Wolfe if he thought that language excluding trends would better the bill. *See H.R. 5055 Hearing*, *supra* note 8, at 6 (statement of Rep. Bob Goodlatte).

415. S. 3728, 111th Cong. § 2(a)(2) (2010).

language raises the standard of similarity that has appeared in previous bills and also eliminates some of the vagueness that has concerned Congress in the past.[416] This language is a compromise between the "closely and substantially similar" and "virtually identical" tests discussed during the H.R. 5055 congressional hearings[417] and is similar to the "substantially indistinguishable" standard used for counterfeit trademarks.[418] This standard does not require that the two garments be so alike that they are impossible to tell apart, just that one could be mistaken for the other.[419]

iii. Proposed Compromise

For the House and Senate to come together on their respective versions of the bill, they will need to compromise on the infringement standard. The "strikingly similar"[420] test has been used in copyright infringement analysis and carries a higher burden than the "substantially similar" test,[421] but it is a lower threshold than "substantially identical."[422] A "strikingly similar" test would be a middle ground between the proposed infringement standards and may be easier for courts to apply because of its use in copyright law. Moreover, the "strikingly similar" standard would be more effective at prohibiting those who intentionally copy designs because the test's function is to show that "access may be inferred if the similarities between the plaintiff's and the defendant's works are so striking that there is no reasonable possibility that the two works were independently created."[423] Some commentators have advocated a line-by-line test for infringement.[424] However, limiting infringing articles to only those that are line-by-line replicas would go too far, as it could allow fashion pirates to skirt the law by making only slight changes to the garment.[425] Such a high standard could render the bill nearly ineffective.

---

416. *See H.R. 5055 Hearing*, *supra* note 8, at 3 (2006) (statement of Rep. Howard L. Berman).

417. *See id.* at 184 (statement of Susan Scafidi); Scafidi, *IDPPPA*, *supra* note 348.

418. *See* 15 U.S.C. § 1127 (2006); *see also* Scafidi, *IDPPPA*, *supra* note 348.

419. *See* S. 3728, 111th Cong. § 2(a)(2) (2010).

420. *See* 20 AM. JUR. 3D *Proof of Facts* § 25 (1993).

421. *See id.* § 13.

422. *See* S. 3728, 111th Cong. § 2(a)(2) (2010).

423. 20 AM. JUR. 3D *Proof of Facts* § 25 (1993) (citing Selle v. Gibb, 741 F.2d 896 (7th Cir. 1984)); *accord* McNamara, *supra* note 342, at 540.

424. *See H.R. 5055 Hearing*, *supra* note 8, at 7 (statement of Christopher Sprigman).

425. *See id.* at 180 (statement of Susan Scafidi).

c. Secondary liability

Both bills, as introduced, would establish secondary liability for design infringement and provide the same remedies as for direct infringement.[426] When the IDPPPA was moved out of Judiciary Committee and to the Senate, this provision was removed.[427] Secondary liability can occur through both vicarious and contributory infringement.[428] While secondary liability is new to *sui generis* design protection, it is not new to copyright; the doctrines of vicarious liability and contributory liability are already well established in case law.[429] Secondary liability is important in fashion design infringement because those who are doing the physical infringing often are aided by others providing samples or photographs of the designs.[430]

d. Home Sewing Exception— IDPPPA Only

The IDPPPA would establish a home sewing exception that would allow "a person to produce a single copy of a protected design for personal use or for the use of an immediate family member."[431] The exception is not intended to permit "the publication or distribution" of sewing patterns to aid in copying the design.[432] This is essentially a fair-use exception similar to those already established in Copyright Act and the Audio Home Recording Act.[433]

7. Recovery for Infringement—DPPA Only

The DPPA would increase the maximum damages available for infringement under Chapter 13 from "$50,000 or $1 per copy"[434] to "[$]250,000 or $5 per copy."[435] The infringed designer would receive whichever amount was greater, whatever the "the court determines to be just" compensation.[436] However, this award potentially exceeds the current

426. *See* S. 3728, 111th Cong. § 2(e)(3) (2010); H.R. 2196, 111th Cong. § 2(e)(3) (2009).

427. *See* S. 3728 (as reported by S. Comm. On the Judiciary, Dec. 6, 1989), 111th Cong. (2010); *see also* Executive Business Meeting, *supra* note 212, at 52:10.

428. *See* MERGES ET AL., *supra* note 273, at 569–72.

429. *See generally* Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd., 545 U.S. 913 (2005).

430. *See H.R. 5055 Hearing*, *supra* note 8, at 11 (statement of Jeffrey Banks).

431. S. 3728, 111th Cong. § 2(e)(3) (2010).

432. *Id.*

433. *See* 17 U.S.C. §§ 107, 1008 (2006); Scafidi, *IDPPPA*, supra note 348.

434. 17 U.S.C. § 1323(a) (2006).

435. H.R. 2196, 111th Cong. § 2(g) (2009).

436. 17 U.S.C. § 1323(a) (2006).

statutory damages cap of $150,000 for willful infringement under copyright law.[437]

Although there is no apparent reason for compensation for infringed designs to exceed that of other areas of copyright, perhaps it is time to examine the available remedies. The last increase to the copyright damages cap, setting it at $150,000, was enacted in 1999 to comply with the Digital Theft Deterrence and Copyright Damages Improvement Act.[438] Because the DPPA's cap on *sui generis* damages would be higher than the current cap on copyright statutory damages, the IDPPPA's drafters were wise not to include this provision.[439] However, if future amendments are made to the monetary recovery available under the Copyright Act, the recovery for design infringement should be adjusted accordingly.

8. Remedy for Infringement—IDPPPA Only

Under the proposed IDPPPA, a design's owner may commence an infringement suit any time after the design is made public.[440] However, before a case can proceed under the IDPPPA, the plaintiff must meet three pleading requirements.[441] First, the plaintiff must "plead with particularity facts establishing that" the design is eligible for protection, meaning that it is both original and novel.[442] Second, the plaintiff must plead that the defendant's design is "substantially identical" to the plaintiff's protected design.[443] Finally, the burden is on the plaintiff to plead facts that establish that the protected design was sufficiently "available" in both location and duration so that it could be "reasonably inferred from the totality of the surrounding facts and circumstances" that the defendant either saw or knew of its existence.[444] This is intended to discourage litigation, a concern expressed by many opponents of previous bills.[445] Because the likelihood of litigation has concerned opponents, this is an important compromise.[446] However, while the cost of litigation will always be an issue no matter the status of the law, this requirement could disadvantage emerging designers who cannot afford the legal fees required to meet such a high burden of

---

437. *See id.* § 504(c)(2).

438. U.S. COPYRIGHT OFFICE, CIRCULAR 92, COPYRIGHT LAW: CHAPTER 5 COPYRIGHT INFRINGEMENT AND REMEDIES, § 504 n.5, *available at* http://www.copyright.gov/title17/92chap5.html (last visited Jan. 1, 2011).

439. *See* S. 3728, 111th Cong. § 2(g) (2010) (containing no mention of specific recovery amounts like the DPPA).

440. *See* S. 3728, 111th Cong. § 2(g)(1) (2010).

441. *See id.* § 2(g)(2).

442. *See id.* § 2(a)(2), (g)(2); Scafidi, *IDPPPA*, *supra* note 348.

443. *See* S. 3728, 111th Cong. § 2(e)(2), (g)(2) (2010); Fasanella, *supra* note 354.

444. *See* S. 3728, 111th Cong. § 2(g)(2) (2010).

445. *H.R. 5055 Hearing*, *supra* note 8, at 86 (statement of Christopher Sprigman); Scafidi, *IDPPPA*, *supra* note 348.

446. *See H.R. 5055 Hearing*, *supra* note 8, at 86 (statement of Christopher Sprigman).

proof through expert witnesses or protracted discovery before they could even advance to trial. It may even discourage them from attempting such litigation.

9. Effective Date

Both bills would "take effect on the date of [the] enactment," meaning neither would apply retroactively to designs.[447] In other words, if a design is made public before either bill is enacted into law, it would not be eligible for protection.[448] This is important for two reasons. First, it would prevent an influx of retroactive applications and allow the Copyright Office to adjust slowly, and, second, it would avoid unfairly punishing those who acted in ways that were legal before the bill's enactment.

IV. CONCLUSION

As Senator Schumer stated, "One of the great things America still has the lead on over other countries is intellectual property; we come up with the best ideas, we find they are often stolen, and this will protect us in one area where we tend to be the leader."[449] Intellectual property protection is imperative to encourage innovation among our country's artists and scientists. Accordingly, the Copyright Act should be amended to include fashion design.

The DPPA and IDPPPA have advanced far beyond their predecessors by making important compromises to allay concerns of objectors to fashion design protection. However, in an attempt to win support, the IDPPPA has adopted confusing language and has devolved into a complicated bill that could fail to adequately safeguard the very designs it was created to protect. Congress should find a compromise between the DPPA and the IDPPPA that would include the registration requirement and a searchable database, establish a "strikingly similar" test for infringement, remove the trends exception, and provide better clarity to the language of the bills in general.

The IDPPPA's movement to the Senate floor[450] is the furthest any of the design bills have progressed since 2006.[451] Although both bills died before a vote could be held,[452] there is hope. The DPPA had twenty-three cosponsors, more than three times the number H.R. 5055 had just three

---

447. S. 3728, 111th Cong. § 3 (2010); *see* H.R. 2196, 111th Cong. § 3 (2009); *H.R. 5055 Hearing*, *supra* note 8, at 218 (statement of the United States Copyright Office).

448. *See* S. 3728, 111th Cong. § 3 (2010); H.R. 2196, 111th Cong. § 3 (2009); *H.R. 5055 Hearing*, *supra* note 8, at 218 (statement of the United States Copyright Office).

449. Executive Business Meeting, *supra* note 212, at 50:39.

450. Executive Business Meeting, *supra* note 212, at 50:20; Nieder, *supra* note 212.

451. *See* S. 3728, 111th Cong. (2010); H.R. 2196, 111th Cong. (2009); S. 1957, 110th Cong. (2007); H.R. 2033, 110th Cong. (2007); H.R. 5055, 109th Cong. (2006).

452. *See H.R. 2196 GovTrack*, *supra* note 209; *S. 3728 GovTrack*, *supra* note 218.

years ago.[453] Additionally, the IDPPPA had the support of both the Council of Fashion Designers of America, long-time supporters of reform, and the American Apparel and Footwear Association, which has previously opposed new measures for design protection.[454] One key player in the outcome will be the Congressional Apparel Manufacturing and Fashion Business Caucus, which recently gained approval on February 2, 2010, and is now a fighting force on the Hill.[455] While the bills are not yet perfect, it looks as though intellectual property protection for fashion design is on its way to gaining wider support and being adopted, finally fitting the needs of the fashion industry.

453. *See H.R. 2196 GovTrack*, *supra* note 209.

454. *See* Scafidi, *IDPPPA*, *supra* note 348.

455. *111th Congress Congressional Member Organizations (CMOs)*, COMMITTEE ON HOUSE ADMIN., http://cha.house.gov/member_orgs111th.aspx (last visited Jan. 1, 2011).

STRICTLY BUSINESS: A HISTORICAL NARRATIVE AND COMMENTARY ON ROCK AND ROLL BUSINESS PRACTICES

JEFF CARTER*

The music business is a cruel and shallow money trench, a long plastic hallway where thieves and pimps run free, and good men die like dogs. There's also a negative side. – Hunter S. Thompson[1]

I. INTRODUCTION 214
II. INDUSTRY STRUCTURE 216
A. *The Record Company* 216
B. *The Record* 218
C. *Royalties and Accounting Practices* 219
D. *Touring* 221
E. *Music Publishing* 222
III. A HISTORICAL NARRATIVE 224
A. *The Birth of the Record Deal* 224
B. *The Blues* 225
C. *The Blues Had a Baby* 232
D. *Rock and Roll Gets Loud—The British Invasion* 236
E. *Success and Excess—The 70's and 80's* 240
1. Touring 240
2. The Live Album 242
3. Independent Promotion 243
F. *You Say You Want a Revolution—The 90's Decline and Digital Delivery* 245
G. *Panic in the Boardroom—The 21st Century and the 360 Agreement* 246
IV. COMMON CONTRACTUAL POINTS OF CONTENTION 249
A. *A Hard Bargain* 249

---

* J.D., *cum laude*, University of Tennessee College of Law, 2010; Bachelor of Science in Psychology, *summa cum laude*, Middle Tennessee State University, 2007. In his former life, Carter owned a music publishing company, maintained songwriting contracts with music publishing companies in Nashville and Los Angeles, and worked as a recording and touring guitarist for several multi-platinum, award-winning recording artists. This article is dedicated to Fred Carter, Jr., who passed along an unlimited enthusiasm for, and invaluable experience within, the recording industry. Carter would like to thank Professor Becky Jacobs for her enthusiastic guidance in pursuing an article on the topic.

1. *Hunter S. Thompson: in His Own Words*, http://www.guardian.co.uk/books/2005/feb/21/huntersthompson (last visited Jan. 14, 2011).

B. "I'm Signed for How Long?" 251
C. Reversion of Rights to Master Recordings & Copyrights 253
V. THE MUSIC BUSINESS IN THE DIGITAL AGE: AN OPINION 255

I. INTRODUCTION

Chuck Berry once said: "Rock and roll is here to pay!"[2] For the young, up-and-coming musician, songwriter, or music publisher, this may be an eye-opening statement given the typical youthful enthusiasm for the music, rather than the business surrounding it. Though music is the heart of rock and roll, commerce is its lifeblood. Indeed, Berry's poignant observation sheds light on the fact that the recording industry is just that: an industry, with opportunity to earn incredible profits while working in an area that touches the individual consumer on many different levels. However, the music business is also an industry rife with both seat-of-the-pants business dealings and cold, manipulative calculation. In both instances, there is ample opportunity to examine the business practices that are peculiar to the recording industry generally, and rock and roll specifically.

The recording industry has undergone many changes throughout its nearly century-long existence. Observing this long history, it is easy to see the ebbs and flows of popular culture as the evolving tastes of the general public have dictated the evolution of recorded music and the business practices surrounding the form. Of the myriad genres of recorded music—along with the back-stories that engendered their various categorizations—none has influenced the industry as a whole with such an earth-shattering, evolutionary force as has rock and roll. Once a fringe genre associated with America's youth, bohemian circles, and sheer rebellion, rock and roll—and the principles promulgated by the genre—has been the multi-billion dollar, genre-transcending bedrock upon which nearly all subsequent recording industry business dealings have been modeled. Of these dealings, none has had a greater impact than the recording contract, or "record deal." Conversely, the music publishing practices that existed prior to and during the formative years of rock and roll brought existing industry custom to bear on rock and roll. This aided its evolving contractual complexity and eventual legitimization as a viable business pursuit.

In order to understand today's recording industry business practices, recording contracts, and music publishing agreements, one must look to the past and delineate the evolution of those agreements and practices over time and across shifting cultural norms. In so doing, one must cast an eye toward varying bargaining power paradigms between recording artists and the business entities that have contracted with those artists for the use and exploitation of their unique talents and services. The record deal and music publishing agreement can be traced to their early origins with agreements

2. Meeting with Chuck Berry, rock and roll pioneer, St. Louis, MO (Sept. 23, 1998).

such as the 1909 recording contract between Victor Recordings and recording artist Billy Murray.[3] However, the historical aspects of recording industry business practices germane to this discussion emerged several years later when individuals like Alan Lomax began to roam the rural Southern countryside in search of oral folk traditions to be recorded for the Library of Congress.[4]

This is not to discount the immeasurable impact of the Victor/Murray agreement; rather, Lomax unwittingly sparked a cultural revolution that would build quietly in the seedier portions of Americana as a diapered Elvis Presley played on his mother's back porch in Tupelo, Mississippi. Interestingly, Lomax undertook his mission in the name of preserving important aspects of America's unwritten cultural heritage before that heritage disappeared from posterity's grasp—a motivation that would prove a far cry from the sheer greed and plasticity that would plague rock and roll in the decades that followed. To be sure, Lomax could never have known that his efforts would ultimately result in the multi-billion dollar music industry that we know today and its importance in, and shaping of, popular culture across the globe.

Alongside Lomax, there are many key players in the emergence of rock and roll as both a form and viable business. The names of legends are common: Muddy Waters, Chuck Berry, Elvis Presley, The Beatles. However, the rest of the story lies with the men behind the music—Leonard and Phil Chess, Sam Phillips, Jerry Wexler, Ahmet Ertegun, and Morris Levy. One would also be remiss to leave out such heavyweight dealmakers as Jay Cooper, Allen Grubman, and the University of Tennessee's own Joel Katz. As was the case with the Chess brothers, Katz entertained no intentions of entering the realm of entertainment; rather, he stumbled into the field after a chance meeting with the Godfather of Soul himself, James Brown.[5] The result of that fortuitous meeting was Katz's successful negotiation of an incredibly favorable contract on Brown's behalf between Brown and his record company.[6] As was the case with the upstart record men who came before him, Katz simply blazed his own trail in the true spirit of rock and roll by negotiating some of the most lucrative agreements known to recording artists at that time and eventually leading the recording industry into the digital music revolution.

Prior to these attorneys' emergence onto the rock and roll scene, recording and music publishing contracts were simple forms thrust across a desk with a take-it-or-leave-it implication that exists to this day. Unfortunately, many of the earliest seminal recording agreements have been lost—some would say conveniently—and later contracts are simply

3. On file with the author.

4. For an in-depth account of Lomax's travels, consult his book. *See* ALAN LOMAX, THE LAND WHERE THE BLUES BEGAN (1993).

5. Press Release, The University of Tennessee (1996).

6. *See id.*

not accessible for public dissemination because of obvious confidentiality issues.[7] However, this paper consults the incredible body of written work and secondary materials as documentation for, and support of, the narrative presented within.

This paper will explore and examine the history of recording industry business practices as created and experienced by the seminal artists, managers, and record industry executives who blazed the genre's collective trail. Further, this paper will discuss various aspects of recording industry contracts and their impact on the recording artist and record company. This discussion will begin with an overview of the recording industry, music publishing, and live performance. This discussion will continue with a historical narrative on the evolution of recording contracts beginning with the oral agreements found during Lomax's historic sojourn through America's countryside and ending with today's complex Multiple Services agreements. Finally, this discussion will conclude with a commentary on particular contractual points of contention, the digital music delivery revolution, and an opinion on the impact of each of these on the future of the recording industry.

II. INDUSTRY STRUCTURE

In an effort to make the topic more easily digestible for the uninitiated, it is worthwhile to give a basic overview of the music industry's structure, in regard to the record company, the record itself, royalties and record company accounting practices, touring, and music publishing. To say that the following discussion will outline the basics would be an overstatement, as it is merely an introductory overview. Accordingly, those readers who seek an in-depth recitation of the ins-and-outs of the music business should seek and consult the many volumes that have been written on the topic.[8]

A. The Record Company

Known as "the labels" in industry circles, major and independent record companies are the typical starting point for making records.[9] In a sense, one can think of the record company as a content aggregation entity because the company seeks to earn profit through the acquisition, reproduction, distribution, sale, and retention of the rights to individual works of recorded

---

7. Author's experience in music publishing.

8. *See, e.g.*, M. WILLIAM KRASILOVSKY & SIDNEY SHEMEL, THIS BUSINESS OF MUSIC 4 (10th ed. 2007); DONALD S. PASSMAN, ALL YOU NEED TO KNOW ABOUT THE MUSIC BUSINESS (6th ed. 2006).

9. Author's experience in music publishing. While one may simply purchase recording equipment and begin the recording process, this was not always the case. Rather, the record companies provided the substantial funding necessary for recording albums. *Id.*

music.[10] Artists sign recording contracts with these entities and, in return, the record company provides the considerable financing necessary for the artist to record his or her record.[11] Though that is the general premise behind the record label/artist relationship, that relationship goes far deeper.

The record company manufactures and ships the recorded product to its distributors, who act as the wholesalers that sell the artist's records to the brick-and-mortar music stores, who in turn sell the music to the public.[12] These distributors also distribute the music digitally through Internet outlets such as iTunes and Rhapsody.[13] The most important function, however, is the record company's use of its vast resources to advertise, promote, and market the music as a product in order to build the artist as a brand.[14] In rare instances, record companies may enter into a licensing agreement with an artist who has chosen to record his or her record independently, with the company acting only as a distribution and marketing service.[15] While this situation is not standard, it is becoming an interesting option for artists who can absorb recording costs and who have built an audience on their own.

As with any business, record companies are split into divisions, which include Sales, Marketing, Promotion, Business and Legal Affairs, and Finance.[16] Unique to the record companies is the "A & R," or Artist and Repertoire staff, who discover, nurture, and sign new talent as dictated by the tastes and culture of the record company.[17] Recently, record companies have begun to add new areas to their operations—in an effort to boost narrowing profit margins while competing for market share in the digital age.[18] In many cases, record companies have begun to add agency, management, and merchandising personnel to their business model, or at least contract for a percentage of the artist's earnings in these areas.[19] These contracts are known as Multiple Services or "360 degree" record deals.

---

10. Telephone Interview with Rob Finan, Attorney (Feb. 15, 2010).

11. Author's experience in music publishing.

12. *Id.*

13. *Id.*

14. *Id.* The building of such a brand is an exceedingly expensive undertaking in the modern music industry. Unquestionably, the record companies are invaluable for their ability to absorb this expense. *Id.*

15. *Id.* This situation is seldom seen because the record companies must surrender a certain amount of control under licensing-only terms. *Id.*

16. *Id.* With independent record companies, these distinctions are often meaningless as one person may work on many facets of the company's business. *Id.*

17. *Id.* "A & R" executives have traditionally served as talent scouts in the rock and roll genre and typically occupy the most tenuous positions at the record companies. *Id.*

18. *See generally* KRASILOVSKY & SHEMEL, *supra* note 8 (examining the effects of technology in the music industry in Chapter 42).

19. Author's experience in music publishing. The author first encountered this while negotiating his own recording contract with H2E/Warner Bros. Records. *Id.*

B. The Record

Most people think of a "record" as the physical album, tape, disc, or downloaded content that is enjoyed on home stereos, car stereos, and iPods. Unknown to most, however, the term "record" as defined in modern recording contracts means both audio-only and audio-visual devices such as videotapes and DVDs.[20] This concept of a record also means "any other device now or hereafter known that is capable of transmitting sound alone, or sound with visual images."[21] Further, current recording contracts define records as "any kind of delivery of music for consumer use, whether sound alone or with visuals," which is meant to encompass all electronic and Internet musical transmissions, regardless of the source of the delivery itself.[22] These definitions of the term "record" are evidence of the record companies' anticipation of new and emerging technologies. This is necessary because of the nature and duration of their recording contracts and the potential for new revenue streams that typically result from new listening and delivery formats.[23] Indeed, the absence of such defining language could have deprived the record companies of millions of dollars in revenue following the widespread emergence of digital downloading.[24]

Similarly, the record companies stand to earn enormous profits every time a new technology arrives, as was the case with the shift from cassette tape to compact disc.[25] In such instances, consumers are certain to replace their beloved music collections in order to accommodate new and ostensibly improved media. When this occurs, the record companies see pure profit, with physical record reproduction and new packaging as the only limited expense.[26] This is because consumers are paying for something that they have already paid for once, resulting in a product in which the record company is required to invest no money for its usual marketing, advertising, or radio campaigns because it already absorbed those expenses upon the record's initial release in the previous format. Thus, the record companies make absolutely certain to account for emerging technological eventualities when contracting with their recording artists.[27]

---

20. PASSMAN, *supra* note 8, at 65–66.

21. *Id.* at 66 (emphasis omitted).

22. *Id.* (emphasis omitted).

23. Author's experience in music publishing.

24. *Id. See generally* KRASILOVSKY & SHEMEL, *supra* note 8 (examining the effects of technology in the music industry in Chapter 42).

25. Telephone Interview with Chris Keaton, Executive, Criterion Music (May 4, 2010). Criterion Music, a music publisher founded in the 1930's, has played a role in the careers of recording artists such as Jackson Brown and Lyle Lovett and owns such copyrights as *Tiny Bubbles* and the musical content of *South Pacific*.

26. *Id.*

27. *Id.*

C. Royalties and Accounting Practices

Record company accounting practices and royalty computations are notorious for their complexity, perhaps purposefully so.[28] The system is set up to ensure that the record company pays nothing to the artist unless and until the company has been reimbursed for each dollar expended on the artist; this includes: recording costs, all promotion and marketing expenses, and any advances received by the artist upon signing.[29] These expenses are recoupable expenses, and it is standard industry knowledge that the artist does not get paid until he or she has "recouped."[30] Under this system, any advance received by the artist is likely the only payment the artist ever receives from the record company because of the industry's standard, accepted accounting procedures described below.[31] Unfortunately for most artists, the only true determination of whether the artist has legitimately recouped is accomplished through an independent audit of the record company's books.[32] However, because independent audits often cost upward of forty to sixty thousand dollars, they are a cost-prohibitive resource for all but the most successful artists who actually stand to benefit should any discrepancies be discovered.[33]

Historically, record companies employed a complicated and tedious system of computing artist royalties that was based on pre-digital delivery systems such as vinyl albums and audiocassette tapes.[34] With the advent of digital music, compact discs emerged and were initially quite expensive to produce, thus generating less profit for both the artist and the record company.[35] However, as the digital format caught on, consumers began to replace their existing vinyl and cassette collections with the higher fidelity compact disc.[36] At the same time, the cost of manufacturing CDs dropped sharply, though the retail price did not reflect the shift for years afterward, resulting in incredible profits for the record companies through the 1990s.[37]

---

28. PASSMAN, *supra* note 8, at 72 (explaining accounting practices).

29. Telephone Interview with Steve Williams, Former Executive, Capitol Records (May 4, 2010).

30. Author's experience in music publishing.

31. *Id.* In many cases, recording artists are not concerned with this situation because they are aware that the real value of their recording contract may lie in the album's ability to drive up music publishing and live touring royalties. *Id.*

32. Telephone Interview with Steve Williams, *supra* note 29.

33. *See* Brian Mencher, *On the Brink of Change: An Examination of the Music Industry's Business Practices*, http://www.bmlawgroup.com/New%20York%20Bar%20EASL%20Article.pdf.

34. *See* PASSMAN, *supra* note 8, at 72.

35. Author's experience in music publishing.

36. *Id.*

37. Telephone Interview with Chris Keaton, *supra* note 25.

The old royalty computations were based largely on paying the artist as little of his or her royalties as possible.[38] This was accomplished under the guise of the suggested retail list price, or SRLP.[39] Here, the record company set a certain SRLP and then automatically deducted a packaging charge from that price, usually amounting to twenty-five percent, an amount which was customarily far greater than the actual cost of packaging.[40] The remaining seventy-five percent became the artist's royalty base, against which the artist applied their recoupment costs and other deductions to arrive at his or her share of the record's profit.[41]

However, the record company was not yet finished reducing the artist's earnings. In addition to the record's actual marketing and promotion budget, the record company added a fifteen-percent "free goods" deduction to the SRLP.[42] Free goods included the records that were given—free of charge as promotional items—to retailers, radio programmers, and concert promoters.[43] While this made a certain amount of sense in breaking an unproven act, the procedure was, and remains, applicable to every record ever released by the artist. Under the old free goods practice, if an artist sold 100,000 records at ten dollars per record, with a ten percent royalty rate, subject to a twenty-five percent royalty base reduction for packaging and a fifteen percent reduction of records actually sold under the free goods practice, the artist could expect to earn only $63,750.00 in pre-tax dollars.[44]

Adding further insult to the artists' injury was the record companies' age-old practice of paying the artist on only ninety percent of net sales in order to account for "breakage."[45] Because records were once made of shellac and were thus breakable, the record companies retained ten percent to cover this expense, a practice which persisted into the modern era when records were no longer breakable.[46] These computations changed in 2006 when, as a result of the digital music revolution, the record companies shifted to a more straightforward system of computing royalties that did away with such outdated practices as breakage, though retaining free goods as a promotional tool.[47] Regardless of that change, many regard the record companies' accounting practices as unfavorable at best and fraudulent at worst; one commentator went as far as to say that the record companies' "prowess and conniving make Enron look like amateur hour."[48]

---

38. Telephone Interview with Steve Williams, *supra* note 29.
39. PASSMAN, *supra* note 8, at 72.
40. *Id.* at 73.
41. *Id.*
42. *Id.* at 73–75.
43. *Id.* at 74.
44. *Id.* at 74–75; *see also* KRASILOVSKY & SHEMEL, *supra* note 8, at 19–21.
45. PASSMAN, *supra* note 8, at 77.
46. *Id.*
47. *Id.*
48. Edna Gundersen, *Rights Issue Rocks the Music World*, USA TODAY, Sept. 16,

Today, artists are paid on a percentage of the wholesale price of the album, resulting in a royalty rate that is often as high as ten percent.[49] Though the free and promotional goods practices persist as accepted methods of marketing both the record and the artist, the newer method of calculation takes into account the idea of "reserves." Records are sold on a one-hundred-percent return basis, which means that the retailer may return to the record company every single record that remains unsold after a set period of time.[50] A typical reserve rate is thirty-five percent, which means that the artist will only be paid initially on his or her ten percent royalty rate as applied to sixty-five percent of the number of records shipped rather than being paid on records actually sold.[51] The remaining reserve royalties are eventually paid if the records are sold, which is known as a liquidation of the reserve.[52] The good news for the artist is that, as his or her notoriety and resulting bargaining power grow, a lower reserve rate can usually be negotiated to the artist's satisfaction, thus resulting in larger initial royalty payments to the artist.[53]

D. Touring

For most recording artists, live performance revenues can be a significant portion of the profits created by the artist's records. Touring is an age-old method by which the artist can accomplish several important and necessary career goals. First, touring is an opportunity to promote the artist's latest product, with most tours nearly always accompanying the release of the artist's most recent recording.[54] This serves as a means of making the artist's fan base aware of the new recording, while at the same time reminding the concertgoer of the artist's past records.[55] Indeed, it is no coincidence that an artist will see a local boost in sales of his or her entire catalog in the days following a live performance in a particular geographic area.[56] This further cements the artist's standing in the fan's view by allowing the fan to experience the music in a more personal, unique, and meaningful way and thus serves to more thoroughly brand the artist.

2002, at 1D.

49. PASSMAN, *supra* note 8, at 69.

50. *Id.* at 71.

51. *Id.* at 71–72.

52. *Id.* at 71.

53. *Id.* at 72. While this may be true, few artists see such a dramatic increase in bargaining power. Author's experience in music publishing.

54. Author's experience in music publishing.

55. *Id.*

56. Telephone Interview with Deana Carter, Recording Artist, Capitol, RCA, and Vanguard record labels (May 6, 2010). It was a common occurrence to see significant increases in local record sales following Carter's concert appearances. *Id.*

Second, and most importantly, touring can be an artist's most lucrative source of income. Because of the nature of record company accounting practices described above, low- and mid-level artists rarely see any profit on record sales, usually because of poor royalty rates or recoupment costs that are rarely fully reimbursed to the record company.[57] Indeed, only top-tier artists usually earn any notable income through their record sales simply because these artists sell a sufficient number of records to reimburse the record company under recoupment and earn a profit.[58] However, nearly every artist has the potential to see a profit from live performances, and many artists support themselves entirely through touring.[59] Further, live shows are the premier venue for selling the artist's merchandise, which can be a significant source of revenue.[60] Indeed, when planned properly, a live tour often becomes the artist's best friend.

E. Music Publishing

Music publishing is the business of copyright ownership.[61] Generally speaking, music publishing companies own, in whole or in part, the copyrights to a songwriter's original compositions.[62] Under music publishing agreements and copyright law, the music publisher retains exclusive rights to the song, including the rights to reproduction, distribution, airplay, and derivative songs.[63] This means that the publisher, not the songwriter, is entitled to make all business decisions regarding the song's use.[64] As a business entity, the music publisher's main function is to take care of the business side of songwriting. Historically, music publishers found artists to record the songs to which the publisher held the copyright, while at the same time turning the song into sheet music and searching for film, television, and stage use opportunities, all in return for a fifty-percent publishing interest in the song.[65] However, as times changed and the artists began to compose their own material rather than looking to professional songwriters for songs, music publishers became something of an investment operation in certain cases. In such cases, the music publisher may sign an

57. *Id.* This is particularly problematic for newer artists who typically incur huge debt with their record labels for promotion costs. *Id.*

58. *Id.*

59. *Id.* In the years following her initial success and high-volume sales figures, Carter earned significant revenues through concert appearances. *Id.*

60. *Id.* In some instances, an artist may accept lower paying concert dates because of the revenue that can be generated via merchandising. *Id.*

61. Author's experience in music publishing. Music publishing would be a meaningless business pursuit without copyright ownership. *Id.*

62. *Id.*

63. *Id.* It is industry custom that these rights are laid out in every music publishing agreement; several examples of music publishing agreements are on file with the author. *Id.*

64. Telephone Interview with Chris Keaton, *supra* note 25.

65. *Id.*

artist/songwriter and pay them a small yearly salary as an advance against potential future royalties.[66] In return, the publisher receives the artist/songwriter's interest in the copyright as a bet that the writer will be successful and yield profitable dividends in the form of future music publishing royalties.[67] In essence, the artist's salary can be thought of as a loan that does not require repayment in the event that the artist never becomes successful.

In regard to revenues, music publishing is a two-way revenue stream, whereby a song's earnings are split with half going to the songwriter and half to the publisher.[68] Put simply, if a song earns one dollar, then the songwriter and publisher each receive fifty cents. Sometimes a song has multiple writers and publishers, which generally results in further equal splits among the parties involved unless contracted otherwise.[69] Under other circumstances, particularly in rock and roll, the recording artist is the songwriter *and* publisher and retains the rights to his or her compositions simply because there is no need to solicit outside recording artists to record the writer/artist's material.[70]

Music publishers and songwriters are paid through professional performance-rights societies such as Broadcast Music, Inc. ("BMI"), The American Society of Composers, Authors, and Publishers ("ASCAP"), and the Society of European Stage Authors and Publishers ("SESAC").[71] These entities monitor public performances across the globe and contract with public broadcast entities such as radio and television broadcasters for payment on the song's use.[72] However, the most important characteristic of music publishing is the music publisher's proprietary interest in the original compositions.[73] These interests exist as perhaps the most important revenue stream that a recording artist/songwriter may enjoy during his or her recording career. Indeed, copyright ownership and music publishing interests have been the source of some of the largest lawsuits ever filed in entertainment.[74]

---

66. Author's experience in music publishing. As a songwriter and recording artist, the author experienced this firsthand through a publishing agreement with Criterion Music Publishing, who signed the author with the hopes of garnering half of the copyright value of the author's then-forthcoming album on H2E/Warner Bros. Records. *Id.*

67. *Id.*

68. Telephone Interview with Chris Keaton, *supra* note 25.

69. *Id.* In some cases, shares of the song may be attributed according to the respective contributions of the co-writers. *Id.*

70. *Id.*

71. Author's experience in music publishing. The author has been affiliated with ASCAP for over a decade. *Id.*

72. Telephone Interview with Chris Keaton, *supra* note 25.

73. *See* E-mail from attorney T.D. Ruth to author (Jan. 30, 2010) (on file with author).

74. Telephone Interview with Chris Keaton, *supra* note 25. Notable suits include a dispute between The Beatles and their 1970's music publisher ABCKO; George Harrison's controversy with Bright Tunes Music; and more recently, a suit between Universal Music

As is the case with recording agreements, agreements between music publishers and songwriters have changed drastically over the years. Songwriter Albert E. Brumley, writer of "I'll Fly Away," signed a two-paragraph agreement[75] in 1936 that is still the source of litigation between Brumley's heirs and the owners of the song's copyright.[76] In its brevity, the agreement failed to detail whether the song was written as a work for hire under the Copyright Act of 1909 or whether Brumley assigned the song's rights to the publisher.[77] The Brumley litigation is notable because "I'll Fly Away" is allegedly one of the most recorded songs in the history of recorded music, second only to John Lennon's and Paul McCartney's "Yesterday."[78]

If a work is made for hire, the employer or other person for whom the work was prepared is deemed to be both author and owner of the copyright unless there has been a written agreement to the contrary, signed by both parties.[79] If in Brumley's case there was an assignment of the rights to the song, Brumley's heirs would be entitled to terminate the assignee's rights to the song and recapture a portion of the royalties under the 1909 Act; if a work for hire, the publisher would retain ownership.[80] It is worth noting that Brumley's heirs have received no earnings whatsoever from the copyright, while the song's publisher has earned millions over the decades.[81] Because of such situations, music publishing agreements, like recording agreements, have become extremely detailed contracts that leave little to the imagination when contemplating potential disputes between the parties to the agreement.

III. A HISTORICAL NARRATIVE

A. The Birth of the Record Deal

Rock and roll did not bring about the birth of the record deal, nor did its precursor—The Blues. Rather, the first rock and roll "record men" adopted portions of existing jazz and pop contracts to fit their particular needs. Accordingly, one cannot explore the business of rock and roll without making note of a key turning point in recording contracts that took place nearly a half-century before Jackie Brenston hit the airwaves with "Rocket 88," the song universally considered to be the first rock and roll song; in 1909, singer William "Billy" Murray contracted with Victor Recordings,

---

Group and Myspace. *Id.*

75. On file with the author.

76. *See* Email from T. D. Ruth, *supra* note 73.

77. *Id.*

78. *Id.*

79. *See* Copyright Act of 1909, §24 (1909).

80. *Id.*; e-mail from T. D. Ruth, *supra* note 73.

81. E-mail from T. D. Ruth, *supra* note 73.

the precursor to the famous RCA Victor corporation, for his services as a recording artist.[82] This contract is notable for the fact that the agreement made no provision for royalties on Murray's behalf as would become the industry standard; rather, the agreement called for Murray to be paid on a yearly retainer of $1,750 in monthly installments of $145.90.[83] In return, Murray provided four recordings per month for Victor, a figure decidedly different from modern recording agreements in that, during Murray's time, a "recording" meant a recorded song rather than the albums for which modern recording artists contract, generally at a pace of one record per year.[84]

A decade later, Murray had become a star and signed a new agreement with Victor.[85] The new agreement was notable for being the first recording contract to offer a recording artist a royalty on records sold in addition to a yearly salary.[86] Modern recording contracts have since done away with artist salaries, instead using cash advances and applying those advances against the artist's later royalty payments, if any, under recoupment.[87] Under this new agreement with Victor, Murray received three-quarters of one cent for every record sold,[88] an amount considerably less than the rate that would be received by the blues and early rock and roll artists of the following two decades. However, the Victor agreement compensated for the financial discrepancy with Victor's inclusion of the $15,000-per-year salary guarantee.[89] Also notable under this agreement was Murray's assignment of his likeness to Victor for marketing and promotional purposes, a practice that persists today and is indeed an indispensable part of breaking a new artist in any musical genre.[90]

B. The Blues

It's not show friends. It's show business. – Sam Phillips[91]

Today's recording and music publishing contracts are tediously long, multifaceted agreements that would be unrecognizable to the blues artists

82. Allan Sutton, *Billy Murray's Victor Recording Contracts: A Case Study, 2009*, http://www.mainspringpress.com/murray_contracts.html (last visited Jan. 14, 2011) (discussing singer William "Billy" Murray).

83. *Id.*

84. *Id.*

85. *Id.*

86. *Id.* Artist royalties were unheard of during that era. *Id.*

87. *Id.*

88. *Id.*

89. *Id.* In present value terms, this would have been a sizeable figure for the time.

90. *Id.* Without assignment of the artist's likeness for visual promotion tools, record labels could simply not function in the modern era. *Id.*

91. Interview with Fred Carter Jr., Record Executive/Musician, in Nashville, Tenn. (Feb. 15, 2010). Carter stated that, first and foremost, Philips was a businessman. *Id.*

who unwittingly gave birth to rock and roll and the behind-the-scenes glut of legal wrangling that is part and parcel of any such profitable business. Often, there were no written contracts at all between the seminal blues record labels and their legendary recording artists; indeed, oral rather than written agreements were the norm.[92] However, as time passed and disputes arose among the parties to these agreements, written contracts took hold and evolved from stock form agreements to the intricately detailed agreements used today.[93] In order to better understand this progression, one must put the discussion into context by examining the parties involved in creating the rock and roll genre and its particular business practices.

Though not the only entity responsible for popular music at the time, Chess Records and its impressive roster of legendary artists, has been cited time and again by both rock and roll royalty and cultural historians as the taking-off point of the music—and a business—that would shape a generation.[94] Chess Records was an independent record company owned and operated by Jewish immigrant brothers Phil and Leonard Chess.[95] Singling out Chess records is by no means an exclusion of other important independent record companies such as Atlantic, Sun, or Roulette; but the fact remains that the Chess brothers were responsible for the single most important cog in the wheel of rock and roll: the recording of an amplified guitar on the blues record "I Can't Be Satisfied" by Muddy Waters.[96] Without that single recording session, there would have been no "Rocket 88" by Jackie Brenston. Applying the theory behind the "butterfly effect,"[97] without "Rocket 88" there would have been no Chuck Berry, thus no Elvis, no Jerry Lee Lewis, no Bob Dylan, no Beatles, no Rolling Stones, no Jimi Hendrix, no Led Zeppelin, no Peter Frampton, no Bruce Springsteen, no Tom Petty, no U2, no Guns N' Roses, no Metallica, no Nirvana, no Coldplay, and ultimately no Kings of Leon. Indeed, electric guitar is the most common thread in the fabric of rock and roll, and it is impossible to imagine the genre's existence without it.

It would appear that, as the old saying goes, history is made by the first-timers. Indeed, the Chess brothers knew very little about business generally

---

92. NADINE COHODAS, SPINNING BLUES INTO GOLD: THE CHESS BROTHERS AND THE LEGENDARY CHESS RECORDS 3 (2000).

93. Interview with Fred Carter Jr., *supra* note 91.

94. *See e.g.*, RICH COHEN, MACHERS AND ROCKERS: CHESS RECORDS AND THE BUSINESS OF ROCK AND ROLL (2004); COHODAS, *supra* note 92; ROBERT GORDON, CAN'T BE SATISFIED: THE LIFE AND TIMES OF MUDDY WATERS (2003) (citing Chess Records).

95. *See* COHODAS, *supra* note 92.

96. COHEN, *supra* note 94, at 15.

97. The so-called "butterfly effect" is a theory based on the notion that important events can begin with the flap of a butterfly's wings, more technically characterized as "sensitive dependence on initial conditions." *See* THE BUTTERFLY EFFECT, http://crossgroup.caltech.edu/chaos_new/Lorenz.html (last visited Oct. 1, 2010).

and absolutely nothing about the music business in particular.[98] In fact, the brothers operated Chess for over two years as a default partnership before bothering to draw up any papers and eventually incorporating the company.[99] The brothers did, however, know what the record buying public wanted and gave it to them in the form of Muddy Waters, Howlin' Wolf, Chuck Berry, Bo Diddley, and many others.[100]

The Chess brothers' business model was simple: find talent, pay the artist a few dollars for his time on a recording session, distribute the resulting record from the back of their Oldsmobiles, and seal it all with a handshake agreement to pay the artist a few pennies on the dollar as a royalty.[101] And pennies were what Chess paid, as the standard arrangement called for a two to three-cent per record royalty rate, a rate that was half that being paid in the pop and big band world.[102] But as business picked up, along with the Chess brothers' knack for earning maximum profits, that royalty payment was usually expressed in the form of a new Cadillac or Oldsmobile delivered to the artist's driveway, a situation leading many music historians to refer to Chess as "Cadillac Records."[103]

Why would the Chess brothers and their artists conduct their business in such a haphazard manner? First, the Chess brothers made up the business as they went along.[104] Neither was educated beyond high school, and neither knew anything about the record business; they ran their affairs with such abandon that Warner Brothers President Joe Smith later called them "bandits."[105] Second, Chess and its artists were each completely dependent upon the other, a situation that led to an inordinately tight bond between them.[106] Without the raw talent streaming to Chicago from the Mississippi delta, the Chess brothers had nothing to sell; without the Chess brothers, that raw talent had no one to whom that talent might be sold.

Third, the lack of written agreements was self-serving for both parties. With no specified contractual methods of accounting, the Chess brothers were free to handle both the company and artists' money in any way they saw fit.[107] Leonard Chess was likely to pay one artist from another artist's account as a means of keeping balance among his artists' finances while promoting harmony given the jealous nature of the artists' competing

98. COHODAS, *supra* note 92, at 56.

99. *Id.*

100. *Id.*

101. Interview with Fred Carter Jr., *supra* note 91. Carter had firsthand knowledge of the situation through his work with Muddy Waters and Ronnie Hawkins & the Hawks. *Id.*

102. COHEN, *supra* note 94, at 149.

103. Interview with Fred Carter Jr., *supra* note 91.

104. COHODAS, *supra* note 92, at 56.

105. COHEN, *supra* note 94, at 67.

106. GORDON, *supra* note 94, at 193.

107. COHODAS, *supra* note 92, at 170.

interests.[108] Likewise, with no contractual clause prohibiting the artists from recording for rival record companies, many artists—most famously John Lee Hooker of "Boogie Chillen" fame—simply hired themselves out to record for other companies under assumed names for quick cash when they needed it.[109] Written contracts were so scorned that one artist remarked: "contracts [did not] mean nothin' . . . [w]e would play for anybody who gave us twenty-five dollars."[110] Fourth, both the Chess brothers and their recording artists were so excited by the prospect of actually making a living out of music that neither cared about, nor paid attention to, business particulars in the beginning.[111]

Finally, and most importantly, was the nature of the artists' collective backgrounds. Almost without exception, the blues pioneers came from the hardscrabble existence known to millions of black sharecroppers during the first half of the twentieth century. Indeed, the grandfather of rock and roll and Chess' biggest star, Muddy Waters, literally stepped off a tractor in Stovall, Mississippi, packed a bag and his guitar, and caught a ride to Chicago to seek a better life.[112] By "better," one must remember that music had yet to explode into a corporate industry accessible to an illiterate southern black man; thus "better" meant anything that did not involve farming cotton twelve hours a day, six days a week, in ninety-eight degree weather.[113]

This was compounded by the sharecropper mentality peculiar to the blues pioneers, known as the "furnish."[114] The furnish was a practice by which the white plantation owners literally furnished everything the black farmhand might need, such as food, clothing, housing, and equipment, in return for the farmhand's services in the cotton fields.[115] Indeed, people in the industry view the relationship that existed between Muddy Waters and the Chess brothers as nothing more than another furnish, whereby in return for Waters' recordings, Chess saw to it that Waters had a new car in the drive, his bills paid in full, and food in his refrigerator.[116] This was apparently the case with other independent labels as well; upon returning

---

108. *Id.*

109. *Id.* at 67; *see also*, WILLIE DIXON & DON SNOWDEN, I AM THE BLUES: THE WILLIE DIXON STORY 118 (1990).

110. COHODAS, *supra* note 92, at 65.

111. This was a common theme among industry pioneers, whose hardscrabble backgrounds lent an initial disregard to specificity of business matters. Interview with Fred Carter Jr., *supra* note 91.

112. *See generally* GORDON, *supra* note 94, at 64 ("Muddy got off [the tractor] . . . walked away from there, and I didn't see him until he come back with his band from Chicago.").

113. Interview with Fred Carter Jr., *supra* note 91. Carter understood this sentiment perfectly, having grown up on a cotton farm in the rural Louisiana Delta. *Id.*

114. GORDON, *supra* note 94, at 8–9.

115. *Id.*

116. *Id.* at 193.

home from an extended tour, James Brown returned to find that King Records' Syd Nathan had purchased Brown a new Cadillac, a new suit, and a case of wine to enjoy during Brown's two-week holiday, only to charge the entirety of the cost for these items back to Brown's account as against Brown's sales royalties.[117] In an interview, Phil Chess asserted that such gifts were not actually gifts because, had they not been considered to be expenses, companies like Chess would have gone out of business.[118]

By all accounts, this was probably far more than Muddy Waters himself ever expected from a career in music. As a younger man, Waters received twenty dollars from Alan Lomax, after the fact, to record two songs in Waters' Mississippi shack during Lomax's epic tenure with the Library of Congress.[119] Interestingly, the initial oral agreement between the two men was sealed with a whiskey as consideration.[120] However, Phil Chess spoke of such "payments" as new cars and liquor as being advances against future royalties and opined that, in the end, the blues artists preferred this treatment.[121] Ever in fear of ruining such a relationship, it is alleged that Waters steadfastedly refused to so much as request a royalty statement from Chess during his lifetime.[122] Further, Waters allegedly refused to challenge Chess' publishing arm, Arc Music Inc., over an agreement that Waters signed which retroactively covered the preceding twenty-five years of Waters' career and encompassed the songs he composed during that period.[123]

Waters was almost completely illiterate and, though the payment for the retroactive agreement likely alerted him to other transgressions by Chess, Waters simply signed the document, took the check, and asked no questions.[124] Longtime Waters associate, the legendary Jimmy Rogers, quoted Waters as saying: "It's a wild-goose chase there with Chess," in regard to receiving any actual royalties as payment.[125] Waters himself denied such allegations later in life, asserting that Leonard Chess was the only reason he had a musical career at all.[126] Eventually, however, Waters filed suit against Arc Music and won an out-of-court settlement based on the aforementioned work for hire portion of the Copyright Act of 1909.[127]

---

117. COHEN, *supra* note 94, at 151.

118. *Id.* at 152.

119. GORDON, *supra* note 94, at xv.

120. *Id.* at 37.

121. COHODAS, *supra* note 92, at 96.

122. GORDON, *supra* note 94, at 226.

123. *Id.*

124. *Id.*

125. *Id.* at 96.

126. COHODAS, *supra* note 92, at 40.

127. DIXON & SNOWDEN, *supra* note 109, at 189. Scott Cameron stated that the basis of Waters' claim was the fact that Arc Music failed to pay Waters a salary for his services as a writer, a factor in the determination as to whether a writer worked as an employee for hire under the 1909 Act. *Id.*

Waters was not the only artist on the receiving end of the Chess publishing inequity. As was customary for the time, nearly every Chess artist signed suspect publishing agreements with Arc Music, effectively giving up half of the earnings on their original compositions.[128] Even worse, many artists were forced to share their portion of the writer's royalty in return for airplay.[129] Famously, Chuck Berry examined his copyright notice for his hit song "Maybellene," only to find out that visionary Cleveland deejay Alan Freed was also listed as a writer.[130] Freed's name appeared as co-writer on many such landmark recordings in return for his agreement to play, and thus promote, the records on his wildly popular and influential radio show.[131] In a bitter twist of irony, Freed—the very man who coined the term "rock and roll"—died penniless following the Payola scandal that emerged from such questionable practices.[132]

In any event, the practice of crediting non-writers was justified as a go-along-to-get-along situation without which the artist's records might never have been heard beyond the walls of the local juke joint.[133] Of course, inaccurate reporting of songwriter credits also operated in the writer's favor in such instances where writers wrote under pseudonyms in order to retain the rights to their songs while under contract to music publishers.[134] Other such unscrupulous business practices began to prevail at that time, such as the infamous 1950's Payola scandal that led to Alan Freed's downfall, whereby deejays were paid cash to play certain records, a topic which is worth an entire study unto itself.[135] However unscrupulous, the Chess brothers likely held the opinion that such tactics were simply the cost of doing business.

---

128. COHEN, *supra* note 94, at 147–48.

129. *Id.* at 163; *see also* DIXON & SNOWDEN, *supra* note 109, at 185 (regarding Chuck Berry).

130. *See* DIXON & SNOWDEN, *supra* note 109, at 185.

131. ED WARD, ET AL., ROCK OF AGES: THE ROLLING STONE HISTORY OF ROCK & ROLL 113 (1986).

132. *Id.* at 245; *see also* THE HISTORY OF ROCK, http://www.history-of-rock.com/payola.htm (last visited Jan. 14, 2011).

133. Interview with Fred Carter Jr., *supra* note 91. Though a distasteful part of the recording business, many artists simply felt that they had no choice in the matter. *Id.*

134. DIXON & SNOWDEN, *supra* note 109, at 114, 118. The most famous writer of the blues, Willie Dixon, claimed many of his compositions under assumed names in order to keep the songs out of the reach of Arc Music, the music publisher with whom Dixon was under contract as a songwriter. *Id.*

135. The Payola issue reared its head again in 2005 when New York Attorney General Eliot Spitzer brought charges against Sony-BMG for modern-day payola practices that included bribes to radio programmers such as vacation packages and other gifts. Settling out of court, Sony-BMG agreed to pay $10 million to charity. Spitzer brought similar charges against Universal Music Group, Warner Music Group, and EMI Music, with each company settling out of court for several million dollars each. *See* KRASILOVSKY & SHEMEL, *supra* note 8, at 383–84.

Other disturbing legal issues were presented by these dealings such as the total lack of legal representation for Waters and other Chess artists. So great was the mindset of the furnish and the fear of ruining their comparatively prosperous relationship with Chess, the artists simply signed the agreements thrust upon them. Most artists signed the agreements without reading them—as mentioned previously, some of these artists were in fact illiterate—and signed them without any explanation by counsel of the artists' rights or duties under the agreement.[136] The issue smacks of duress as well, though it is well settled that duress is not present so long as alternatives are available to the parties, regardless of the distastefulness or limited nature of those alternatives.[137] Indeed, it is no violation of general contracting principles to drive a hard bargain as did the Chess brothers, and the simple fact remains that the artists signed the agreements freely and certainly could have sought counsel even at the risk of losing the deal with Chess.

Finally, many Chess artists allegedly signed various recording and music publishing agreements while under the influence of alcohol.[138] By many accounts, Leonard Chess was famous for arranging office meetings with his artists during which it was understood that important contractual and business matters would be discussed.[139] Prior to the agreed upon meeting time, Chess would set out several bottles of his artists' favorite liquor along with drink mixers, drinking glasses, and ice cubes.[140] Upon the artist's arrival, Chess' secretary would inform the artist that Chess had to step out for an emergency and that they should make themselves at home in his office.[141] Chess would then disappear and have his secretary call him after the waiting artists had had time to become intoxicated, after which he would reappear, offer the men another drink, and produce whatever documents that required the drunk artist's signature.[142]

---

136. COHODAS, *supra* note 92, at 116.

137. WILLISTON ON CONTRACTS §71:43 (4th ed. 2003).

> A contract may be unenforceable due to unequal bargaining power if the party challenging the contract has been the victim of fraud or overreaching; if the party was coerced into signing; if it was denied the opportunity to seek legal advice prior to entering into the agreement; or if the contract is somehow unconscionable. Merely taking advantage of another's financial difficulty is not duress where the difficulty is not the result of the actions of the advantaged party.

Id.

138. DIXON & SNOWDEN, *supra* note 109, at 100.

139. *Id.*

140. *Id.*

141. *Id.*

142. *Id.*

Such shenanigans remained part of the industry for years and continue to this day.[143] During the so-called "peace and love" of the Sixties, prominent rock and roll manager Barry Friedman was cut out of a lucrative deal between himself and Buffalo Springfield.[144] As a means of persuading Friedman to release the band, he was given a large amount of marijuana and then driven around in a limousine for hours with instructions for the driver not to stop under any circumstances for food or drink until Friedman signed the document releasing Buffalo Springfield from their contract.[145]

Because intoxication is a basic formation defense under contract law,[146] the Chess artists might have had any unfavorable or one-sided agreements signed while intoxicated declared null and void under general contract law, had they so desired.[147] As the years progressed, however, the artists became aware of the effects of such unscrupulous business practices on their careers. At the same time, these artists became aware of their popularity and the profits generated by it, thus leading to inquiries into financial statements and a multitude of resulting lawsuits which were settled out of court and thus do not exist as part of the public record.[148] Ultimately, all of these factors coalesced into the appearance of detailed agreements in the blues and the emerging rock and roll genre and the disappearance of certain, though certainly not all, distasteful business practices.

C. The Blues Had a Baby

Muddy Waters sang: "The Blues had a baby, and they named it rock and roll."[149] Though merely a song lyric to Waters, the statement was a perfect description of the new genre born of the blues to poor, southern,

---

143. Author's experience in music publishing and live touring. A favorite ploy of club owners is to provide oblivious acts with prodigious amounts of alcohol before the show in order to circumvent complaints about pay and accommodation. *Id.*

144. FRED GOODMAN, THE MANSION ON THE HILL 65 (1997).

145. *Id.*

146. The Restatement (Second) of Contracts states:

> A person incurs only voidable contractual duties by entering into a transaction if the other party has reason to know that by reason of intoxication: (a) he is unable to understand in a reasonable manner the nature and consequences of the transaction, or (b) he is unable to act in a reasonable manner in relation to the transaction.

RESTATEMENT (SECOND) OF CONTRACTS § 16 (1979).

147. *Id.*

148. Muddy Waters, Howlin' Wolf (a.k.a. Chester Burnett), and others initiated such suits against Chess Records and Arc Music Corp. and eventually settled out of court for undisclosed amounts. DIXON & SNOWDEN, *supra* note 109, at 188–89.

149. MUDDY WATERS, *The Blues Had a Baby and They Named it Rock and Roll*, on HARD AGAIN (1977).

white boys in Sam Phillips' Sun Records recording studio. Just as Chess Records cornered the market on the blues, Phillips assembled an impressive stable of artists whose names are synonymous with rock and roll: Roy Orbison, Carl Perkins, Jerry Lee Lewis, and Elvis Presley.[150] Like Chess, Sun was an independent record company run by a man with everything to lose. However, Sam Phillips and his Sun label proved to be the lynchpin that first tied the blues to rock and roll and then introduced rock and roll to the corporate world in the form of a young dynamo born in Tupelo, Mississippi and bred in Memphis, Tennessee.

Both popular culture and the recording industry took a quantum leap forward when Elvis Presley changed the world with his 1954 classic, "That's Alright Mama."[151] Written by blues artist Arthur Crudup, Presley's recording marked the explosive intersection of black music and the white, teenage audience that would catapult rock and roll into a legitimate business pursuit, a pursuit that would compel the corporate world to take increasing notice of the form. Sun Records owner and Presley's first producer, Sam Philips, famously stated: "If I could find a white man with the Negro sound and the Negro feel, I could make a billion dollars."[152] Of course, the recording industry as a whole grossed only $205 million in the previous year,[153] a figure that would see fifteen-percent-per-year increases in the decades to come thanks to the mass appeal of rock and roll and the changes in consumer post-war, entertainment-based discretionary spending that coincided with the emergence of the form.[154] Indeed, Phillips' statement was indicative of both his creative and business acumen and demonstrated his sense of the untapped market.

Unfortunately for Phillips, such prescient thought was insufficient leverage for catapulting his then-current successes into long-range profit. In a move made out of sheer financial necessity, Phillips sold Presley's contract to RCA Victor in the largest music industry signing of the day.[155] Phillips has asserted that he did not regret the move in light of the fact that a wildly successful record such as "That's Alright Mama" could actually prove to be a financial disaster for an independent record company like Sun.[156] In such situations, a small record company is forced to wait for months for cash flow to appear on record sales, and thus the company cannot keep pace with retail orders due to the lack of funds necessary to press and ship the large quantities of records required when one breaks out

150. Interview with Fred Carter Jr., *supra* note 91. Carter recorded and performed with Orbison and Lewis, among others. *Id.*

151. WARD ET AL., *supra* note 131, at 80.

152. *Id.* at 77.

153. *Id.*

154. Author's experience in music publishing.

155. WARD ET AL., *supra* note 131, at 112.

156. *Id.* at 117. The Beatles later had the same effect on Vee Jay Records that Elvis had on Sun Records. *Id.*

as a hit.[157] As a result, the independent record company owner could literally see his company starved to death by its own success.[158]

In the RCA deal, Phillips received $35,000 for Presley's Sun Records contract along with the rights to any previously recorded material, plus a $5,000 bonus upon Presley's signing of a new RCA Victor recording contract.[159] Additionally, RCA affiliate Hill & Range Music established a music publishing company in Presley's name that would give Presley a fifty-percent stake in any song that Presley recorded for a five-year period.[160] The RCA deal was notable for the facts that the dollar amount involved was unheard of for the time and that Presley was neither a songwriter nor a music publisher.[161] Further, Presley received a royalty rate of five cents per record, an increase of two cents over his previous Sun royalty rate of three cents per record sold,[162] the standard established by other independents such as Chess records during the previous decade.

Of course, RCA could afford to take the risk on Presley because it was a corporate entity whose music division could be supported in times of economic distress by its many other divisions.[163] This phenomenon would recur in later decades as major corporations—both in and outside of the music business—began to successfully gamble corporate funds on rock and roll.[164] RCA's gamble on Presley paid off handsomely, as Presley accounted for over fifty percent of the company's total record sales within three months of signing his recording contract.[165] Later in Presley's career, RCA gambled again and purchased the rights to all of Presley's future royalties in return for a $5.4 million lump-sum payment.[166]

The RCA deal was a testament to the keen business savvy of Presley's new personal manager, "Colonel" Tom Parker and to the bargaining power created by Presley's immense popularity. Parker was a shrewd businessman who learned his particular brand of dealing while working in the circus[167] and whose methods would be copied by leading rock and roll managers

---

157. *Id.*

158. *Id.*

159. *Id.* at 112. Presley's RCA contract would prove even more lucrative as his success grew. Parker later renegotiated the RCA agreement for then unheard of terms. PETER GURALNICK, LAST TRAIN TO MEMPHIS, 353–54 (1994).

160. WARD ET AL., *supra* note 131, at 112; *see also* GURALNICK, *supra* note 159, at 232.

161. In accomplishing this feat of negotiating, Elvis' manager, "Colonel" Tom Parker, set the precedent for the music business manager's garnering of every penny possible for his or her client. Interview with Fred Carter Jr., *supra* note 91.

162. GURALNICK, *supra* note 159, at 231–32.

163. WARD ET AL., *supra* note 131, at 111. At the time, RCA owned such brands as Whirlpool, whose refrigeration units were selling quite well. *Id.*

164. Author's experience in music publishing.

165. WARD ET AL., *supra* note 131, at 119.

166. STAN SOOCHER, THEY FOUGHT THE LAW: ROCK MUSIC GOES TO COURT 3 (1999).

167. GURALNICK, *supra* note 159, at 165–66.

during the decades following Presley's breakout.[168] For his outsized services, Parker charged a heavy price, taking twenty-five percent of Presley's gross earnings,[169] a figure ten percent higher than the industry standard fifteen percent management rate.[170] As had the blues artists that he so admired, Presley signed both his agreement with Parker and his RCA agreement without the advice of legal counsel, a common theme running throughout the history of rock and roll. Answering allegations that Parker advised Presley to not seek legal advice, Parker defended himself by stating that if Presley "wanted it, he got it."[171]

Allegedly, Parker would have Presley do anything in the name of the almighty dollar, and it has been alleged that Presley never opposed Parker in any of his business endeavors, if Presley even knew the details of those endeavors.[172] In a typical moment of insulating Presley from his own business dealings, Parker snatched a live performance agreement from Milton Berle's hands when Berle moved to place the contract in Presley's hand just prior to Presley's appearance on the Milton Berle Show.[173] Clearly, Parker had no desire to share business details with Presley. In another instance demonstrating the psychological power of management over the artist, Parker allegedly told Presley: "If you ever do anything to make me ashamed of you, you're through."[174] It is not known whether this statement was the result of Parker's psychological control over Presley, the result of Parker having covered up Presley's early sexual exploits and thus maintaining Presley's image, or a combination of both.

Most notable in Parker's handling of the Elvis phenomenon was his pioneering introduction of rock and roll merchandising as a legitimate source of artist wealth and promotion. For the first time in history, a merchandising campaign was aimed directly at the teenage population, whereby Presley's face was plastered on every type of consumer item imaginable and sold through such outlets as Sears, Montgomery Ward, and Woolworth's.[175] Parker employed his carnie background with flair, earning himself and Presley almost one million dollars during 1957, representing the customary five percent licensing royalty on $18 million in wholesale sales.[176] This success would later be emulated and improved upon by rock managers following in Parker's considerable footsteps.

---

168. Author's experience in music publishing.

169. GURALNICK, *supra* note 159, at 258.

170. Author's experience in music publishing.

171. SOOCHER, *supra* note 166, at 7.

172. Interview with Fred Carter Jr., *supra* note 91. Carter stated that Elvis would not entertain any talk of business whatsoever without Parker's direct involvement in the conversation. *Id.*

173. GURALNICK, *supra* note 159, at 262.

174. *Id.* at 283.

175. *Id.* at 354.

176. *Id.*

However, not all of Parker's for-profit practices were completely legitimate. A particularly beloved Parker scheme involved his squeezing of every imaginable profit from Presley's live performances. During the hours prior to show time at Presley's early concerts, Parker would have someone place five-by-seven note cards and golf pencils in the seats of a particular venue with instructions for the audience members to provide personal information including name, address, and phone number, under the guise of providing that information to Presley's fan club.[177] However, Parker allegedly sold this information to direct marketing firms for "on the side" profit rather than using it for any fan club purposes—and purportedly without Presley's knowledge—a decidedly unethical use of personal information to say the least.[178] Such opinion aside, it is an unquestioned fact that Parker was a genius at marketing Presley, and both men reaped incredible profits as a result.

D. Rock and Roll Gets Loud—The British Invasion

By the early 1960s, rock and roll appeared to be drawing its final breath. A fatal airplane crash had claimed the lives of rock and rollers Buddy Holly, Ritchie Valens, and The Big Bopper,[179] Jerry Lee Lewis had married his thirteen-year-old cousin and in so doing had enraged the public,[180] Chuck Berry had been sentenced to federal prison for a violation of the Mann Act,[181] and Elvis had returned from the Army more interested in filming movies than in shaking, rattling, or rolling.[182] However, salvation—and a huge shot in the recording industry's collective arm—was about to board a plane out of London bound for New York City. The Beatles landed in America in 1964 and infused the record industry with an energy not seen since Elvis appeared nearly a decade prior.[183]

Here, the business of rock and roll began to move at a far faster pace, with the standards in music, contracting, and merchandising solidly established under the aforementioned influence of Leonard Chess, Sam Phillips, Elvis Presley, and Colonel Tom Parker. Indeed, later acts simply capitalized on what their forebears created, with the central changes coming in the form of increasingly elevated record sales, record company profits, and artist advances and royalties.[184] Put simply, rock and roll had become a

---

177. Interview with Fred Carter Jr., *supra* note 91. Carter stated that Parker probably learned such tactics while employed by the circus early in his career. *Id.*

178. *Id.*

179. WARD ET AL., *supra* note 131, at 192–95.

180. *Id.* at 178.

181. *Id.* at 208.

182. *Id.* at 211, 245.

183. ALBERT GOLDMAN, THE LIVES OF JOHN LENNON 152–53 (William Morrow & Co. 1988).

184. Interview with Fred Carter Jr., *supra* note 91. Carter stated: "They weren't doing it

cash cow, a cow that would fatten itself over the next four decades before reaching a legitimate slump after the digital revolution. But in 1964, the world was the rock and roll band's oyster and, for the first time, it came *to* America rather than *from* it.

It is at this point that the story of rock and roll business practices necessarily steps onto the fast track. It is fair to say that the emergence of the Beatles and, later, the Rolling Stones, Led Zeppelin, Jimi Hendrix, and others, though culturally and musically earth-shattering, brought about few changes in the actual business structure of rock and roll. Indeed, it appears that these artists and their managers walked through doors opened by Elvis Presley and Colonel Tom Parker in regard to contracting and the new multi-million dollar rock and roll merchandising industry.[185] Just as Presley had rounded his rough edges at Parker's behest and in the name of commercialization almost a decade earlier, the Beatles transformed from hard-edged rockers into a teddy bear merchandising phenomenon.[186] Regardless, the British bands brought about America's complete immersion in rock—the "and roll" portion having gone by the wayside with their arrival—as the most influential entertainment form of the 1960s, during which time the corporate entertainment entities truly began to take an active role in the business of rock.[187]

Among their many accomplishments was the Beatles' use of film as a means of promotion. While Elvis had successfully navigated the waters of Hollywood and thus promoted himself, the Beatles used film to promote their music.[188] With *A Hard Day's Night*, the Fab Four preempted MTV by almost twenty years.[189] Interestingly, the movie was actually born as a business ploy that would allow United Artists to acquire the rights to a Beatles album.[190] Capitol Records, the Beatles' American record company, initially refused to release the Beatles' records in America because the company felt the music to be unsuitable for American audiences.[191] Anticipating the Beatles' eventual success in America, however, United Artists signed the group to a three-picture deal that included soundtrack albums, thus garnering United Artists the rights to the album version of *A Hard Day's Night*, which sold quite well.[192] The interesting result of this plan was that American mothers and fathers accompanied their children to

---

different[ly], they just did it better." *Id.*

185. *Id.*

186. GOLDMAN, *supra* note 183, at 163. In a carefully planned publicity shot, photographer Harry Benson staged The Beatles in a pillow fight, a fact which disgusted John Lennon and magnified his distaste for becoming "what every little girl most wanted." *Id.*

187. GOODMAN, *supra* note 144, at xi.

188. *See, e.g.*, A HARD DAY'S NIGHT (United Artists 1964).

189. BARRY MILES, MANY YEARS FROM NOW 160 (1997).

190. GOLDMAN, *supra* note 183, at 165.

191. *Id.*

192. *Id.*; *see also* MILES, *supra* note 189, at 157–58.

theaters across the country much as they had for Elvis Presley's Blue Hawaii,[193] thus cementing the widespread commercial appeal that would earn the group both mass adulation and massive profit.

However successful the group became, they were nonetheless victims to the poor business decisions of their manager, Brian Epstein. According to an Arthur Young audit, the group had earned $154 million by 1968, yet found themselves in dire financial straits.[194] The Beatles later discovered that they were going broke because they had subjected themselves to Epstein's faulty business decisions in regard to their initial contract with EMI Records and Dick James Music Publishing.[195] Under the EMI recording contract, the group saw the low royalty rates typical for unproven talent during that era which garnered them one cent per single record sold, a rate that persisted through the group's early lucrative years.[196] Falling prey to his English civility, Epstein had simply refused to renegotiate the Beatles' contract after the group became successful because he felt it to be in bad taste to go back on one's word.[197]

Even more interesting is the fact that, after that contract expired in 1965, the Beatles operated for over a year and a half without a recording contract with EMI because Epstein was simply unable to arrive at what he felt to be the best terms for the group.[198] Epstein ultimately negotiated a somewhat better deal for the Beatles, garnering a generous royalty increase and advance payment against future profits; however, in so doing Epstein committed the group to EMI for another ten years.[199] Further, Epstein had committed the group's original compositions to Dick James Music Publishing in much the same manner as Leonard Chess had committed his stable of artists to music publishing agreements with Arc Music two decades prior, effectively depriving Lennon and McCartney of a portion of their music publishing income.[200]

Under the Dick James agreement, James himself retained a fifty-percent interest in every Lennon-McCartney composition as administrative publisher while Lennon and McCartney split the remaining half between themselves and Epstein, effectively cutting Epstein into the Beatles' publishing business.[201] As had Elvis before them, the Beatles entered into both the Dick James agreement and the Epstein management agreement without advice of counsel, once again raising ethical issues between

193. GOLDMAN, *supra* note 183, at 170.
194. *Id.* at 327.
195. *Id.* at 331–34.
196. *Id.* at 332.
197. *Id.*
198. *Id.*
199. *Id.*
200. *Id.* at 332–33.
201. *Id.*

themselves and their manager.[202] McCartney has asserted that he and Lennon had no idea that a song had ownership rights and that the two writers simply signed the contract that was put in front of them with no understanding of the underlying issues related to copyrights or music publishing, much like the blues artists before them.[203] Though such practices had once been the norm, managers and music publishers became increasingly unable to dupe their artists because of the high profile nature of the disagreements in rock that differed greatly from the relative obscurity under which the early blues independents operated.[204] From a business perspective, such stories as the Lennon/McCartney example are worth noting as cautionary tales, warning even the most successful artists of the pitfalls associated with recording industry contracts.

However, not all of the Beatles' business dealings ended on a sour note. In an effort to plan around England's draconian tax laws, the Beatles incorporated their business pursuits under the umbrella of Apple Corps, which exists as a profitable business entity to this day.[205] In a move that proved nearly as profitable as anything the group accomplished musically, the Beatles copyrighted the Apple name worldwide.[206] Thus, when Apple Computers was formed years later, the computer company was forced to negotiate a lucrative agreement with Apple Corps regarding trade usage of the Apple name.[207] However, that agreement specified that Apple Computers could only use the Apple name for so long as the company did not pursue music in any way.[208] Thus, Apple Computers' foray into the world of digital music with its iTunes application, store, and iPod mp3 players forced the computer company to again negotiate an agreement with Apple Corps, resulting in an even more lucrative agreement for Apple Corps.[209]

Apparently learning from Apple Corps' current successes and the Beatles' past mistakes, McCartney later acquired the music publishing rights to the entire Buddy Holly publishing catalog, a move that eventually earned McCartney millions in addition to his own songwriting and publishing royalties.[210] Indeed, the Dick James publishing agreement taught

---

202. MILES, *supra* note 192, at 145.

203. *Id.* at 146. McCartney further asserted that the Beatles were on such a creative roll that it would be impossible to take a year off to sort out their business troubles, if in fact such troubles even existed. It would appear that, like most artists, the Beatles knew very little about their business affairs at all, evidenced by the fact that upon dissolution of the Beatles' partnership, the members did not even know a partnership agreement existed. *Id.* at 147, 578.

204. Author's experience in music publishing.

205. MILES, *supra* note 192, at 440.

206. *Id.* at 581.

207. *Id.* at 581–82.

208. *Id.*

209. *Id.*

210. TIMOTHY WHITE, ROCK LIVES 126 (1990).

McCartney the true value of copyright ownership and music publishing, a lesson McCartney shared with Michael Jackson when the two collaborated as songwriting partners during the 1980s.[211] In a shrewd business move, Jackson heeded McCartney's advice and promptly went behind McCartney's back and purchased the copyrights to the entire Lennon-McCartney publishing catalog.[212]

E. Success and Excess—The 70's and 80's

1. Touring

By the end of the 1960s, the Beatles had fallen out of love with one another,[213] and youth culture had earnestly embraced the louder, grittier musical styles embodied by such acts as Jimi Hendrix, The Rolling Stones, The Who, Cream, and Led Zeppelin. While these groups enjoyed wild success and eventual wealth from record sales, their prime contribution to the rock business model was through extended touring. Live touring can be the financial bedrock and saving grace of the mid-level artist's career and a major source of the top-tier artist's wealth and fame. Prior to the emergence of the big rock acts of the 1970's, artists simply chartered a bus and hit the open road as a means to promote their records and earn a bit of extra money in the clubs and at county fairs.

Later, rock acts changed the status quo by turning their live performances into enormously profitable spectacles experienced by the masses in stadiums across the nation. Indeed, by the early 1970s, the top rock acts were grossing in excess of $100,000 per night.[214] This boom in touring began with the first rock festivals, including Altamont and Woodstock, which proved the existence of a nearly insatiable market for large concert events. However, the most important factor in the success of the live rock concert was the re-characterization of the acts themselves by emerging talent agents. In an illustrative quip, top industry executive Clive Davis asserted that as soon as the performers began being called artists, their fees instantly shot up.[215]

Rock agent extraordinaire Frank Barsalona founded the agency, Premier Talent, [216] which would play a vital role in the transformation of rock touring. While employed with talent agency General Artists

---

211. *Id.* at 148–50. McCartney asserts that Jackson sought out advice as to how he might invest his *Thriller* profits wisely; Jackson promptly began licensing the Lennon-McCartney songs for all manner of marketing campaigns. *Id.*

212. *Id.* at 148–49.

213. *Id.* at 125.

214. Telephone Interview with Carol Peters, Manager of recording artist Heart and former Motley Crue management employee (Jan. 18, 2010).

215. GOODMAN, *supra* note 144, at 24.

216. *Id.* at 25.

Corporation during the 1960s, Barsalona was assigned the acts for whom no other agents cared to do business: the rock bands.[217] Barsalona stated: "Motion pictures, television were the hip place to be. In those days rock was this bastardized part of show business that was going to be over in a couple of years. There was no future, no talent."[218] Observing the success of the Beatles' first trip across the Atlantic, however, Barsalona saw the future of music and left General Artists to found Premier Talent, an agency that specialized in rock touring.[219] Barsalona's instincts proved infallible and Premier became a wild success, eventually grossing $50 million in 1977 alone, ten percent of which went directly into Barsalona's pocket, as is customary for booking agents in all musical genres.[220] Indeed, Barsalona's success was indicative of rock's overall success during the 1970s.

Another major stepping-stone in the growing success of rock touring was the introduction of corporate sponsorship. As rock record sales exploded and the resulting tours became more lucrative, corporate sponsors began to see the tours as a useful means of gaining market share with the emerging youth consumer base. In one of the most profitable instances in rock, Budweiser and other major retailers sponsored the Rolling Stones' "Steel Wheels" tour, which grossed over $70 million.[221] Such sponsorships are important to the artist because the involvement of a sponsor can significantly defray tour expenses and, in instances involving the largest corporate sponsors, absorb those expenses completely.[222]

Today, touring promises less financial opportunity than it once did, particularly for low and mid-level artists. This is due to several factors, including the glut of acts performing, the stranglehold of now-corporate talent agencies on the independent local promoter, and a new trend known as the "buy-on." The idea behind the buy-on is that a new and unproven act will pay a major headlining artist a sum of money for the opportunity to tour with the headlining artist as its opening act.[223] With so many new artists seeking to fill so few available slots, the buy-on has become something of an additional marketing expense that causes the new act to see little, if any, profit from any tour requiring the artist to buy on.[224] The

217. STEVE CHAPPLE & REEBEE GAROFALO, ROCK 'N' ROLL IS HERE TO PAY 124 (1977).

218. GOODMAN, *supra* note 144, at 24.

219. *Id.* at 25.

220. CHAPPLE & GAROFALO, *supra* note 217, at 125.

221. Jack Doyle, *Stones Gather Dollars*, THE POP HISTORY DIG (Dec. 3, 2008), http://www.pophistorydig.com/?tag=rolling-stones-steel-wheels-tour. Not to be left out, retailers such as Macy's joined in the sponsorship as well, creating "Rolling Stones boutiques" within their regular retail centers. *Id.*

222. Telephone Interview with Deana Carter, *supra* note 56.

223. PASSMAN, *supra* note 8, at 345.

224. Telephone Interview with Deana Carter, *supra* note 56. Because of this, Carter declined several opportunities to buy onto a particularly popular artist's tours during the late 1990s. *Id.*

present view of this practice depends upon one's perspective: the new artists and older industry insiders see it as a cutthroat means of squeezing every possible profit from an established artist's career by defraying their touring expenses, whereas those associated with the established artist see the buy-on as a proper means of maximizing the utility of their business model.[225] From either view, the buy-on is indicative of the "pay-to-win" mentality that is so prevalent in the recording industry.

Paramount to the profitability of live touring is tour merchandising. Once the sole province of retail and mail order, as conceptualized by Colonel Tom Parker in his marketing of Elvis, rock merchandising eventually took on a life of its own at a rate that paralleled the success of the tour itself. Today, mid-level and top-tier artists can, on an average live performance, expect to earn a substantial profit from the sale of t-shirts, posters, and all manner of items bearing the artist's name and likeness.[226] Many artists contract with third-party merchandisers to provide their services, an agreement under which the artist receives a percentage of net sales; in other cases, more successful artists choose to maintain in-house merchandising in order to retain all profits generated.[227] It is no secret that top-tier artists can gross in excess of $100,000 per live concert on merchandise sales, while mid-level acts may generate upwards of $25,000 per night.[228]

2. The Live Album

The success of the revitalized live performance gave rise to a spate of live recording releases during the 1970s. Once considered commercial failures released only to please the enormous egos common to modern rock stars, the live album found new life in the 1970's with the success of Peter Frampton's *Frampton Comes Alive*.[229] After embarking on a profitable relationship with Frank Barsalona that garnered Frampton ever-increasing concert earnings, Frampton, Barsalona, and Frampton's manager decided to release a live record that would effectively market Frampton by showcasing his raw talent on stage.[230] Selling over twelve million copies and spawning three rock staples,[231] *Alive* became an unqualified success and generated over $50 million in record sales and tour receipts.[232] *Alive* was also notable

---

225. *Id.*

226. *Id.* Carter earned substantial revenues through merchandise sales. *Id.*

227. Author's experience in music publishing and live touring.

228. *Id.*

229. *Id.*

230. GOODMAN, *supra* note 144, at 312.

231. *Id.* The songs "Baby, I Love Your Way," "Show Me the Way," and the guitar classic "Do You Feel Like We Do" each appeared on *Frampton Comes Alive*. PETER FRAMPTON, FRAMPTON COMES ALIVE (1976).

232. GOODMAN, *supra* note 144, at 312.

for the fact that it was a double album and thus required twice the packaging and twice the split among music publishers, both of which have been heavily frowned upon by industry executives.[233] Unfortunately for Frampton, the record's success failed to translate into a maintainable career: he was soon bankrupt because of poor management and unwise personal business decisions.[234] Regardless, *Alive* was such a wild success that it legitimized rock as a viable business pursuit in the eyes of the various corporations.

Following *Frampton Comes Alive*, domestic and multi-national corporations began to purchase American record and music publishing companies. Though not typically thought of as possessing keen business minds, a handful of British rockers followed these transactions with great interest.[235] In the ultimate example of a rock-goes-corporate maneuver, rocker David Bowie issued publicly traded bonds, so-called "Bowie Bonds," that were asset-backed securities funded by, and paid from, the profits earned on his personal record sales.[236] This practice has since been embraced and employed by other artists in the rock genre.

3. Independent Promotion

The emergence of rock as such a huge source of revenue eventually sparked the era of the independent promoter. Existing as one aspect of rock's more suspect business investments, independent promoters were freelance record promoters who wielded an immense amount of power during the 1970s and 1980s. Dubbed the "new payola,"[237] independent promotion was a simple concept: the independent promoters garnered radio airplay for the recordings owned by whatever record company bid the highest for the independent promoters' services.[238] The central contention over independent promotion involved its extortion-like qualities and the often less-than-reputable means by which these individuals were able to ensure that airplay; these methods typically involved such tactics as cash payoffs, drug-induced coercion, and out-and-out personal threats.[239]

---

233. Telephone Interview with Chris Keaton, *supra* note 25. Albums were traditionally limited to ten songs, giving each publisher an even ten-percent interest. Increasing the number of songs per album necessarily reduced each individual music publisher's interest and thus reduced their earnings on the record's sales. *Id.*

234. GOODMAN, *supra* note 144, at 312–15.

235. It was widely known that the British bands such as The Yardbirds, The Beatles, and The Rolling Stones took their craft more seriously than their American counterparts during the early days, possibly because the English groups viewed rock as a way out of the rigid English caste system. *Id.* at 27–28.

236. Daniel Kadlec, *Banking on the Stars*, TIME.COM, http://www.time.com/time/innovators/business/profile_pullman.html (last visited Jan. 14, 2011).

237. FREDRIC DANNEN, HIT MEN 13 (1990).

238. *Id.* at 5.

239. Interview with Fred Carter Jr., *supra* note 91. Carter asserted that there were often

The record companies were willing to engage in the practice in order to maintain their competitive edge in the increasingly cutthroat marketplace in the wake of the 1950s' Payola scandal, despite chances of violating the new Racketeer Influenced and Corrupt Organizations statute ("RICO").[240] Examining the ramifications of RICO on the record industry, one legal commentator observed: "The threat of RICO liability created an incentive for record companies to retain independent contractors for record promotion in order to insulate themselves from imputed criminal liability or complicity."[241] The incentive did not stop there. The independent promoters' true power was not derived from their ability create a hit; rather, it was their ability to prevent a song from being aired that incentivized the record companies to pay their exorbitant fees.[242] RCA President Elliot Goldman remarked that "it wasn't payola, it was extortion—the price you had to pay to be in the business."[243] Adding insult to injury, the independent promoters' respective radio stations did not always add the song in question to the station's rotation, or else placed the song in an unfavorable time slot such as the "lunar" rotation of the 2:00 a.m. hour.[244]

The power of the independent rock promoter is best demonstrated by an incident surrounding the release of Pink Floyd's *The Wall*. By 1983, CBS Records was spending between $8 and $10 million annually on independent promotion,[245] and the industry as a whole was spending thirty percent of its gross earnings on the practice.[246] In an attempt to break the independent promoters' hold on the industry, CBS President Dick Asher persuaded Floyd to kick off its world tour in the key rock market of Los Angeles with the understanding that CBS would not hire any independent promoters to push airplay of the chosen single, *Another Brick in the Wall*.[247] Because other stations across the country had begun airing the song, Asher assumed that at least one of the four major, taste-setting Los Angeles stations would simply follow suit for one of the biggest bands in the world. Asher was wrong.[248]

Angered by Asher's threat to their livelihood and attempted blow to their outsized egos, the promoters flexed their collective muscle and caused

---

no limits as to what might be done to garner radio airplay. *Id.*

240. DANNEN, *supra* note 237, at 14.

241. J. Gregory Sidak & David E. Kronemyer, *The New Payola and the American Record Industry: Transactions Costs and Precautionary Ignorance in Contracts for Illicit Services*, 10 HARV. J. L. & PUB. POL'Y 521, 538 (1987).

242. DANNEN, *supra* note 237, at 13.

243. *Id.* at 16.

244. *Id.* at 17.

245. Sidak & Kronemyer, *supra* note 241, at 535.

246. DANNEN, *supra* note 237, at 15.

247. *Id.* at 10.

248. *Id.*

a complete radio blackout of Pink Floyd across Los Angeles.[249] The promoters' power over radio was so great that the stations refused to play the record even in the face of overwhelming public demand for it, which was astonishing given the hype surrounding the Los Angeles concert and its coinciding with the release of *The Wall*.[250] In order to confirm the results of his experiment, Asher eventually relented and lifted the ban on the independent promoters only to discover that *Another Brick in the Wall* immediately rose to the top rotation slot of every Los Angeles rock station.[251] This result was an incredible blow to CBS's own credibility and influence within the rock radio market.[252]

The source of the promoters' influence over radio was obvious. In one instance, a single DJ admitted to receiving over $100,000 in cash payments as monthly "birthday gifts" from a prominent independent promoter.[253] Describing the situation more specifically, the DJ asserted that for every week that he added another one of the promoter's songs to the on-air rotation, he would receive an additional $500 to $1000 in his birthday card.[254] In an example of a different means by which the independent promoters accomplished their mission, an independent promoter working for Atlantic Records got an on-air disc jockey so high on marijuana that the DJ fell asleep in the studio bathroom.[255] Taking advantage of the situation, the promoter took over the radio show and proceeded to play nothing but new Atlantic Records releases for two hours.[256] Indeed, free drugs were second only to cash payments when it was time to promote a new record, a fact asserted by independent promoter Bob Garcia who stated that the record companies would give deejays "anything from hash brownies to free meals to bottles of booze" to get their records played on the air.[257] Though such practices have been toned down to some extent, record companies continue to employ whatever means necessary to break a record.[258]

F. You Say You Want a Revolution—The 90's Decline and Digital Delivery

The rock industry maintained its stability and prospered throughout the various difficulties posed by the Payola scandal of the 1950s, the drug excesses of the 1960s, and the independent promoters of the 1970s. It is fair

249. *Id.* Upon learning the details of the situation, Pink Floyd's manager became so angry that he personally hired the independent promoters to get the record played. *Id.*

250. *Id.*

251. *Id.* at 10–11.

252. *Id.* at 11.

253. *Id.* at 14.

254. *Id.*

255. GOODMAN, *supra* note 144, at 186.

256. *Id.*

257. *Id.* at 187.

258. Telephone Interview with Deana Carter, *supra* note 56.

to say that the 1980s and early 1990s saw even more growth and stability than was experienced in the decades prior. However, the bubble burst during the late 1990s with the emergence of digital music delivery via the Internet. Volumes have been written on this landmark topic and thus this discussion will remain true to its purpose of examining the historical underpinnings of rock and roll business practices rather than examining every aspect of the emergence of digital downloading, the single most dramatic change in the recording industry since the early 20th century when records supplanted sheet music as the industry's most powerful force.[259] The key fact central to this discussion is that as Internet downloading increased, record sales declined and revenues decreased proportionately.[260]

Record executives blame the situation on online file sharing and other forms of piracy, while detractors fault the record companies themselves. Indeed, the record company business practices of the 1990s were shocking in their excess, and the common record executive attitude during that time was that record sales could never fail the major companies.[261] This attitude led record executives to ignore the impending changes brought on by the Internet, and, by the time there was widespread industry acceptance of those changes, it was far too late to take appropriate action.[262] Thus, as sales declined and record companies began to lay off their employees, the companies were forced to strategize new ways to maintain their bottom lines. The most recent, and most controversial, of those new strategies is the 360 degree record deal.

G. Panic in the Boardroom—The 21st Century and the 360 Agreement

In response to the difficulties presented by the emergence of digital music delivery and the resulting decrease in profits, the record companies have begun to present recording artists with multiple services, or "360," deals.[263] These agreements are so named because of the record companies' complete encircling of each of the artist's revenue streams, revenue which was historically not included as part of the traditional record deal, including touring, merchandising, and music publishing.[264] Under these agreements,

---

259. *Sheet Music*, NEW WORLD ENCYCLOPEDIA (Nov. 6, 2008), http://www.newworldencyclopedia.org/entry/Sheet_music.

260. Author's experience in music publishing. Fault for this may lie with the recording industry for filling albums with undesirable songs as a means to deliver one or two hit singles. Now that consumers can readily purchase their favorite songs individually as digital singles, there no longer exists the need to purchase an entire album that may or may not contain other desirable material. *Id.*

261. *Id.*

262. *Id.*

263. *Id.* The author was presented with such an agreement between himself and H2E/Warner Bros. Records. *Id.*

264. *Id.*

the record companies take a percentage of the artist's earnings in each of these areas as a means to justify their current outdated business model and bolster their bottom lines.[265] A typical example of such a situation may involve the record company contracting with the artist for a percentage of the artist's gross touring revenues, merchandising sales, and music publishing royalties, each of which has historically been kept completely separate from the artist's recording agreement.[266]

Prominent factors leading the record companies to present and justify these deals as the new standard agreement have been the widespread practice of online piracy and the death of the album. The latter was caused by per-song and subscription music services such as iTunes and Rhapsody, each of which has played a substantial role in the decreased record sales that have threatened to ruin the major record labels. In a strange twist, there are instances when the artists themselves undermine their own record sales, and thus harm their record companies, by encouraging concertgoers to record and digitally share the artists' live performances.[267] From the artist's perspective, this interesting twist serves dual purposes. First, this is a means of antagonizing the record company because the artist often feels that he or she has been taken advantage of under his or her recording contract.[268] Second, the practice cultivates an us-against-them bond between the artist and his or her fans as a means to bolster the artist's revenue streams unrelated to record sales if not under 360 degree contractual terms.[269] Dave Matthews Band and the Grateful Dead are famous for this practice.[270]

Critics of these agreements have suggested that the record companies change their outdated business models as a means of dealing with the drastically changed marketplace.[271] Accordingly, most, if not all, record companies have indeed altered their internal practices if not their overall business models. Once famous for their extravagant spending, the record companies have tightened their collective belts by reducing executive salaries, eliminating certain positions while combining others, drastically reducing expense account limits, and generally reducing or altogether eliminating discretionary expenditures associated with the industry.[272] In one instance well known to industry insiders, Capitol Nashville once spent

265. Telephone Interview with Deana Carter, *supra* note 56. Carter has avoided such agreements in her own career. *Id.*

266. *Id.*

267. Telephone Interview with Chris Keaton, *supra* note 25.

268. Author's experience in music publishing.

269. *Id.*

270. Telephone Interview with Chris Keaton, *supra* note 25.

271. *Id.* Indeed, record labels have become more parsimonious in several different areas, but a widespread overhaul of industry practices has yet to occur. *Id.*

272. *Id.*

$500,000 on a single party in celebration of a particular artist's platinum sales award, an expenditure that would certainly not be seen today.[273]

The record companies have countered the criticism of the 360 deal with the contention that the terms are absolutely necessary as a means to remain in business in the face of potentially devastating financial losses resulting from low sales figures.[274] Because the costs of producing records have not decreased in relation to reduced record sales, this argument is not without merit. Indeed, the record companies must retain the services of recording studios, musicians, sound engineers, pressing plants, and advertising firms whether an artist's record sells one copy or ten million. Further, industry insiders argue that, by partnering with an artist in additional facets of his or her career, the record company is better suited to serve the artist as an individual and to promote the artist as a product.[275]

From a legal perspective, 360 deals raise several issues that can only be addressed through careful negotiation, if indeed the artist stands in a position to negotiate at all. Because 360 deals combine a mixed bag of formerly independent stand-alone contracts,[276] the parties must determine their respective obligations, most importantly that of performance and breach. With recording and music publishing contracts, artists could typically rescind the contract only when the record company refused to record and release product.[277] Likewise, an artist would only be released from management and live performance contracts in the event of a similar breach by the artist's personal manager or booking agent.[278] The combining of these services with the recording contract raises the question of how to determine when the record company has in fact breached. However, the most troubling of these issues involves the record companies' acquisition of a percentage of the copyrights to the artist's original songs.[279] Because of this acquisition, the record company reaps the benefits of the agreement between itself and the artist long after the agreement has expired.[280] Because so many previously independent revenue streams are combined under the 360 deal umbrella, the artist's attorney must carefully delineate every obligation under the agreement in order to ensure the artist's protection in the event of breach by the record company and vice versa.

---

273. Author's experience in music publishing.

274. Telephone Interview with Steve Williams, *supra* note 29.

275. *Id.*

276. 360 deals encompass formerly stand-alone interest such as music publishing, touring, and merchandising. Author's experience as a recording artist.

277. PASSMAN, supra note 8, at 109.

278. Author's experience in music publishing.

279. *Id.* Taking part of the artist/songwriter's copyrights deprives him or her of what has historically been a significant source of revenue for that individual. *Id.*

280. *See* Ian Brereton, *Is this the Beginning of a New Age?: The Unconscionability of the "360-Degree" Deal*, 27 CARDOZO ARTS & ENT. L.J. 167, 194 (2009).

IV. COMMON CONTRACTUAL POINTS OF CONTENTION

A. A Hard Bargain

A major point of contention regarding the formation of recording industry contracts centers on the artist's perceived lack of bargaining power in negotiating with the record company. Without question, it is a "take it or leave it" situation for almost every new or unproven recording artist.[281] This is a simple matter of numbers, of supply and demand: singers, musicians, and songwriters stream into Los Angeles, New York, and Nashville by the thousands, hoping to fill a bare handful of available slots in an industry governed by elusive and ever-shifting standards and tastes. Because of this, the record companies and music publishers enjoy an unbelievable amount of bargaining power, typically offering the artists a standard boilerplate deal that heavily favors the company.[282] Faced with no alternative and eager for a chance to succeed, many artists happily sign with record companies under unfavorable and onerous terms, entering into agreements that could be deemed unconscionable based on most artists' lack of any meaningful alternative.[283]

Recording artist Don Henley summed up the situation succinctly, asserting that new acts either do not care about or do not understand the issues, or simply choose not to "rock the boat."[284] In a particularly nasty turn, record companies have even gone so far as to sign these eager newcomers simply as a means to bind the artists and thus prevent them from being signed by another company and later emerging as competition for the signing record company's established acts.[285] Although the artist may eventually rescind the agreement because of the record company's failure to record and release product, the ordeal will have served a legitimate business interest for the record company and at the same time ruined any momentum the artist may have gained toward a successful career.

Under general contract law, one-sided contracts are not necessarily unconscionable, though if the court finds the contract to have been unconscionable at the time it was entered into, the court may refuse to enforce the agreement, may elect to enforce the remainder of the contract without any unconscionable clauses, or may simply limit any unconscionable result.[286] In *Williams v. Walker-Thomas Furniture Co.*,[287]

---

281. *See* Gundersen, *supra* note 48 (quoting singer and songwriter Tom Waits: "[m]ost people are so anxious to record, they'll sign anything").

282. *See* Brereton, *supra* note 280, at 178–79.

283. Author's experience in music publishing.

284. Gundersen, *supra* note 48.

285. Author's experience in music publishing.

286. U.C.C. § 2-302 (2003).

287. 350 F.2d 445 (D.C. Cir. 1965).

the court asserted that one factor determinative of whether a contract is unconscionable is whether the weaker party had any meaningful alternative to the contract at the time the parties entered into the agreement.[288] The *Williams* court concluded that a rent-to-own contract could be unconscionable because of the lack of any meaningful alternative as well as a lack of bargaining power on the claimant's behalf.[289] Though distinguishable on the facts, the same can be said of recording contracts entered into by creatively talented individuals who possess absolutely no bargaining power and face no reasonable or meaningful alternative by which they might earn a living with their particular abilities.[290] Armed with clever attorneys, the record companies prevent the establishment of harmful legal precedent by settling related claims before they reach the public record.[291] Interestingly, no United States court has ever rendered a decision as to whether recording agreements are in fact unconscionable.[292]

Conversely, one could argue that recording artists enjoy a huge shift in bargaining power when they become successful and subsequently renegotiate their contracts; extremely high-profile artists may demand and receive extraordinary amounts of money.[293] Indeed, the record companies could simply refuse to renegotiate and thus force the artist to perform or sue, though such situations rarely occur.[294] Even when the record company capitulates, it does not actually find itself on the losing end of the bargain because, by retaining that artist's services, the record company will continue to earn a measure of profit on whatever terms are negotiated in addition to the exorbitant profits already earned on the artist's previous and current record sales. Further, the record company may enjoy other intangibles by retaining the artist's services such as maintaining a high industry profile for having retained the artist on the company roster, which in turn promotes the record company further as a desirable home for the "next big thing" yet to be discovered.[295]

---

288. *Walker-Thomas Furniture Co.*, 350 F.2d at 449.

289. *Id.* at 449–50. After asserting that an unconscionable contract could be rescinded, the *Williams* court remanded for further findings on the unconscionability issue. *Id.* at 450.

290. *See generally* Brereton, *supra* note 280 (discussing in depth the doctrine of unconscionability and its applicability to recording contracts).

291. Bill Holland, *Artists' Lawyers Debate Contracts*, BILLBOARD, Sept. 29, 2001, at 70–71.

292. *See* Brereton, *supra* note 280, at 176.

293. Notable artists have garnered millions with such renegotiations, including Madonna, Michael Jackson, Janet Jackson, U2, The Dixie Chicks, and others. Author's experience in music publishing.

294. *Id.*

295. *Id.*

B. "I'm Signed for How Long?"

Another major contention between record companies and recording artists is the duration of the recording contract. This issue has been both litigated in state and federal court and addressed in Senate hearings.[296] Generally, seven-year contracts are the industry standard, though that duration is actually connected with the number of records to be delivered by the recording artist.[297] The chief complaint with term lengths is that the contracts typically require the artist to deliver seven albums during that term, a nearly impossible feat with modern touring schedules, lengthy promotional campaigns, and extended radio charting cycles.[298] Thus, though the duration of years may be met under the terms of the agreement, the artist continues to owe the record company product. The result of this is that artists may be forced to remain under contract with record companies far beyond the artist's commercial desirability or else face legal action for breach if they refuse to perform.[299]

However, the courts have decided cases in the artists' favor.[300] In a landmark term length suit between MCA Records and pop singer Olivia Newton-John, MCA sought to enjoin Newton-John from recording for any other record company until she specifically performed under her existing contract for a set number of years.[301] At that time, recording agreements allowed the record company to extend the artist's contract every time an album was delivered late.[302] In *Newton-John*, the plaintiff argued that the agreement should be limited to the number of years actually stated in the agreement, without regard for extensions.[303] Finding in Newton-John's favor, the California Court of Appeals agreed and found that no artist could be enjoined from recording for another company for a period of time longer than which the artist could fully perform under his or her existing contract.[304] Thus, recording agreements have since defined the agreement's duration in terms of the artist's delivery of a certain number of albums within a stipulated number of years.[305]

---

296. *See* KRASILOVSKY & SHEMEL, *supra* note 8, at 15–16 (1997); Laura M. Holson, *Music Stars Complain About Stringent Contracts*, N.Y. TIMES, Sept. 6, 2001, at C1, *available at* http://www.nytimes.com/2001/09/06/business/music-stars-complain-about-stringent-contracts.html.

297. Interview with Fred Carter Jr., *supra* note 91.

298. *Id.*

299. *See* KRASILOVSKY & SHEMEL, *supra* note 8, at 15.

300. *See* PASSMAN, *supra* note 8, at 100–01 (discussing the Newton-John and Martin cases, discussed *infra*).

301. *See* MCA Records, Inc. v. Newton-John, 153 Cal. Rptr. 3d 153, 154 (Cal. Ct. App. 1979).

302. PASSMAN, *supra* note 8, at 100.

303. *See generally id.* (discussing *Newton-John*, 153 Cal. Rptr. 3d, at 155).

304. *Newton-John*, 153 Cal. Rptr. 3d at 155.

305. PASSMAN, *supra* note 8, at 101.

However, other wrinkles in the term lengths remained. Pop crooner Dean Martin was party to an agreement with Warner Brothers Records wherein the term length continued until delivery of all the recordings under the contract.[306] After a six-year period without recording, Martin notified Warner out of the blue of his intent to record and thus receive his agreed-upon payment for recording.[307] Because Martin was, by then, far less popular than he had been upon signing with Warner, the company refused to pay for another Martin recording and was thus forced to settle the disagreement after Martin sued.[308] This situation led the record companies to amend their term length clauses because, under the Martin terms, a recording contract could potentially draw out forever.[309] Demonstrating the other side of the same Warner Brothers contractual clause, progressive rocker Frank Zappa arrived at the Warner offices carrying four albums' worth of material.[310] Zappa declared the recordings to be the remaining product under his contract along with his intention to sign immediately with another record company.[311] This turn of events led the company to further amend its term length by adding that an artist is prohibited from starting a new album until that artist has delivered the preceding album and that the new album cannot be delivered within a six-month period of delivering the preceding album.[312]

Opining in a Billboard Magazine article covering the topic of recording contract term lengths, prominent music attorneys asserted the aforementioned problem of settlements between the record companies and their artists that serve to keep their disputes out of the public record.[313] New York music attorney Wallace Collins asserted that it is absolutely to the record companies' advantage to settle contract disputes between themselves and their artists before trial because, by so doing, they prevent the establishment of potentially unfavorable legal precedents. In Collins's words, this strategy allows the record companies to "make the rules of the game instead of letting the courts make them."[314] Another prominent entertainment attorney, Owen Sloane, asserted that the issue of potentially disadvantageous contract terms is exacerbated by the fact that the courts may be unsympathetic to the often world-famous, multi-millionaire recording artists who bring such claims, deeming such parties to be experienced businesspeople who can fend for themselves.[315]

---

306. *Id.*
307. *Id.*
308. *Id.*
309. *Id.*
310. *Id.*
311. *Id.*
312. *Id.*
313. Holland, *supra* note 291, at 70–71.
314. *Id.* at 70.
315. *Id.*

On the other side of the argument, Recording Industry Association of America attorney Carey Sherman has asserted that lengthy recording contracts are necessary because of the time and money required to develop new artists.[316] This reasoning is based on the fact that many of these artists will never record a successful album, resulting in huge financial losses for record companies.[317] Industry studies suggest that fewer than five percent of signed recording artists become profitable investments for record companies, and for every successful album recorded, they absorb losses in excess of $6.3 million.[318] Accordingly, the record companies incur the bulk of the risk in signing a new or unproven recording artist.[319]

Another defense of the contractual term lengths relates to the aforementioned renegotiation of the artist's recording contract while the artist is at the height of his or her popularity, often resulting in considerable concessions by the companies.[320] Such renegotiation typically binds the artist to a longer term and cross-collateralizes the artist's past and future works for recoupment purposes.[321] On the balance, however, industry insiders have opined that the record companies have begun to reconsider lengthy contractual obligations as a natural result of ongoing pressure from the recording artists, coupled with the uncertainty of the recording industry's future in the digital age.[322] Notable examples of this trend are Universal Music Group and Sony BMG, both of which have announced plans to shorten the duration of their contracts.[323]

C. Reversion of Rights to Master Recordings & Copyrights

Under the recording industry's standard recording agreement, the record company acquires the rights to the artist's master recordings in perpetuity, thus preventing the artist from acquiring the rights to those recordings even though the artist may end up paying for the recordings in full.[324] Though the record company initially invests the required capital to record the artist's record, the artist ultimately pays the entirety of those master production costs through recoupment, whereby the record company withholds the artist's royalties until the company has been fully reimbursed

316. Holson, *supra* note 296.

317. *Id.*

318. Gundersen, *supra* note 48.

319. Holson, *supra* note 296; *see also* Gundersen, *supra* note 48.

320. Interview with Fred Carter Jr., *supra* note 91. These concessions often occur as increased royalty rates. *Id.*

321. Holland, *supra* note 291, at 70.

322. Telephone Interview with Steve Williams, *supra* note 29.

323. KRASILOVSKY & SHEMEL, *supra* note 8, at 459.

324. Telephone Interview with Carol Peters, *supra* note 214 (discussing artist's regaining control of their master recordings).

for the production costs.[325] In many cases, the record company also recoups all promotion and marketing costs under the reasoning that such costs are necessary adjuncts to the production process.[326] However, the more successful artists are able to divide these costs with the record company or even avoid the costs altogether upon renegotiation of their recording contracts.[327] Otherwise, artists may use the reversion of master recordings as a bargaining tool. In a rare case, hard rockers Motley Crue regained control of their Elektra Records master recordings, using the company's unwillingness to pay a sizable amount of money for a one-year option extending the band's contract to the band's advantage in negotiating their outright release from the record company.[328]

More generally, recording artists have been unable to regain control of their master recordings, either through lobbying to have the work-for-hire portion of the copyright law changed or through contractual negotiations with the record company itself.[329] However, the record companies have universally contracted around the work-for-hire issue with the inclusion of contractual provisions requiring the recording artist to assign all rights in the master recordings to the record company in the event the recording is deemed to not be a work-for-hire contract in the first place.[330] Senator Orrin Hatch summed up this predicament perfectly during a Senate hearing on the topic: "[T]he record business is the only industry in which the bank still owns the house after the mortgage is paid."[331] The situation becomes particularly egregious in cases wherein the record company refuses to make the record available for sale, as may be the case with older or less profitable recordings, whereby the record company deprives the artist of both the ownership interest in the master recordings and the ability to earn profit from record sales.

Further complicating the situation are the limitations placed on recording artists via the industry's standard auditing practices. Generally, record companies make royalty payments and issue royalty statements twice per year, but only after the company has recouped all of its expenses from the artist.[332] Once the artist decides to audit the company, certain limitations arise: the artist cannot hire an auditing firm on a contingency basis, the firm cannot be involved in any other audit of the record company,

---

325. Author's experience in music publishing. It is useful to bear in mind that creative accounting practices rarely demonstrate that an artist has in fact recouped, thus keeping the artist in debt with the record company or music publisher. *Id.*

326. *Id.*

327. *Id.*

328. Telephone Interview with Carol Peters, *supra* note 214 (discussing artists' regaining control of their master recordings).

329. *Id.*

330. *Id.*

331. Gundersen, *supra* note 48, at 1D (quoting U.S. Senator Orrin Hatch).

332. *See* Mencher, *supra* note 33.

and, most importantly, the auditor is allowed access only to certain documentation related directly to record sales and excluding such items as manufacturing costs.[333] Finally, the auditor may only review each document once, and the artist is bound to fulfill the recording contract even if it is proven that a royalty underpayment has occurred.[334]

With sweeping changes on the horizon for the industry as a whole, the reversion of rights to master recordings will likely become less of an issue as established recording artists increasingly pay for the production of their own records and subsequently license albums to the record company for mass production, packaging, and distribution. This may allow artists to use the record company's vast resources while retaining ownership of the master recordings and thus retaining control of their artistic visions. Further, such master licensing agreements may allow the artist to avoid the potential pitfalls found within today's 360 agreements. On a smaller scale, lesser known and emerging artists have begun to pay for and market their own records through the Internet, live touring, and other grass-roots promotional campaigns.[335] By so doing, such artists retain complete control of their recorded product and, by retaining a far greater percentage of sales royalties, may make up for whatever profits may be lost by going outside the standard record company promotional machine.

V. THE MUSIC BUSINESS IN THE DIGITAL AGE: AN OPINION

The digital revolution has played an undeniable role in changing the recording industry as we know it. Piracy has become a more pervasive problem than was initially imagined, leading to decreased record company profits and the resulting multiple services contracts. Of course, artistic theft has existed alongside mankind from the beginning; indeed, without a certain amount of "borrowing," as some artists have put it, the historical oral and visual traditions upon which rock and roll is based would never have led to today's entertainment industry. A traditional saying among songwriters is that "good artists create, but great artists steal."[336] This was certainly the case during the days of Chess Records and the subsequent years that saw the development of rock and roll, wherein blues artists took liberties with the songs of other blues artists and Led Zeppelin co-opted various Willie Dixon songs for their own electric exploitation.[337]

---

333. *Id.*

334. *Id.*

335. Author's experience in music publishing. A shining example of this phenomenon is the rise of the pop band OneRepublic, who gained notoriety through exposure on MySpace. *Id.*

336. *Id.*

337. *Wilson & Alroy's Record Reviews*, http://www.warr.org/zep.html (last visited Jan. 14, 2011).

However, with the advent of widespread digital downloading, copying, and file sharing across all musical genres, music theft now occurs by consumers as well as between artists and composers. But to be fair, many modern recording artists encourage such behavior among their fans. As mentioned earlier, it is a well-documented fact that certain acts, most notably the jam bands that are the new rock and roll, regularly encourage their fans to record and distribute their live concerts, radio, television, and film appearances in the name of "spreading the music."[338] In a sense, this could be considered akin to the artists' own endorsement of stealing from their record companies, which by nature of the sales figures of such bands are the true parties in interest to their recording contracts: only the record companies stand to earn any profit from such live performance-oriented bands because of the bands' limited sales figures.

The ramifications of the digital music revolution are significant. Much discussion has taken place regarding the business consequences to both the industry as a whole and the artists as individuals. Interestingly, little has been made of the artistic consequences in music industry circles. What does all this mean for the creative artist who is the driving force behind, and root of, the recording industry? In a word: success. Without question, today's accessibility to music production, distribution, and sales resources has enabled members of the general public to become both artist and business entity, with their success limited only by their particular talents and resources. Without the stress and limitations typically cast upon the artist by the record company, the independent artist could potentially reach new heights in artistry and produce the fresh sounds once associated with rock and roll.

Not long ago, the music industry was closed to those outside "the know" who possessed the limited knowledge, and even more limited financing, required for access to and success within the music business. Previously, an artist needed the record company to pay for the recording, production, distribution, marketing, and sales of his or her product; now the independent artist has become the record company and enjoys digital access to the world of music distribution where such access previously did not exist. The natural result of this is that anyone with a glimmer of talent, appropriate recording software, an Internet connection, and a little ingenuity can become a star. Thus, the law of natural selection could take over and allow only the fittest, or at the very least the most clever or talented, artists to survive.

While this is bad news for those individuals who have undertaken great strides to exist in the music industry as professionals, it is certainly good news for both the consumer and the Kid with Big Dreams who lives in Everytown, USA. Further, the present situation should serve to bring the

338. Telephone Interview with Carol Peters, *supra* note 214.

most talented artists to the fore based on factors that at one time had far less bearing on a new artist's success or failure.

For the business interests, the digital revolution has caused the record companies to partner with other business entities in ways never before seen in the industry. Now it is common for the record companies to seek partnership with the cellular telephone industry as a way to access their ring tone markets as a viable source of profits.[339] Record companies have also ventured into the world of gaming, often choosing to release an established artist's newest recording as part of the soundtrack to the video game itself as a promotional and marketing ploy rather than releasing the song to radio in the traditional manner.[340] This is particularly useful because such use significantly reduces the promotional costs paid by the record companies for traditional methods. Finally, the record companies have forayed into their artists' revenue streams via the aforementioned multiple services agreements as a means to reach any profits that may be available to them.

The result of these practices, which some would characterize as desperate, appears to point toward the downfall of the major record company system as the industry has known it. At the very least, the record companies will be forced to continue rethinking their approach to the market while at the same time opening the lids of their corporate coffers less freely. To be sure, gone are the days of the six-figure expense accounts, extravagant promotional junkets, personal drivers, and the like, once common perks for even the most mediocre record company A & R executive. However, the buying public will always demand new music because it is a human imperative to seek entertainment. As the individual artist/independent music business entity continues to meet these demands with distinctive product and efficient business practices, the record companies will be forced to respond in kind, eventually leveling the playing field and allowing rock and roll to continue evolving.

339. Author's experience in music publishing.

340. *Id.*

TORT LAW—COMPARATIVE FAULT—ORIGINAL TORTFEASOR RULE IN TENNESSEE

Banks v. Elks Club Pride of Tennessee 1102, 301 S.W.3d 214 (Tenn. 2010).

I. INTRODUCTION

In March 2006, plaintiff Alice J. Banks attended a social event at an Elks Lodge and sustained serious injuries to her back when the chair on which she was seated collapsed.[1] The plaintiff consulted Dr. Robert H. Boyce, an orthopedic surgeon, who recommended and performed lumbar surgery.[2] After completing the surgery, Dr. Boyce realized that he had performed a portion of the surgery at the wrong vertebrae.[3] As a result, the plaintiff had to undergo a second surgery.[4] Following the surgeries, the plaintiff was transferred to Cumberland Manor Nursing Home ("Cumberland Manor"), where she developed a serious staphylococcus infection that necessitated additional surgeries as well as extensive care and treatment.[5]

In March 2007, the plaintiff filed suit against the Elks Lodge,[6] alleging that its negligence caused her back injuries.[7] In May 2007, the plaintiff filed a separate suit against Dr. Boyce and his employer, Premier Orthopaedics and Sports Medicine, P.C., alleging medical negligence and medical battery based on Dr. Boyce's performance of an unauthorized procedure.[8] In January 2008, the two cases were transferred to the same court and

---

1. Banks v. Elks Club Pride of Tenn. 1102, 301 S.W.3d 214, 216 (Tenn. 2010). Because a full hearing regarding the facts had not yet taken place when the opinion was written, the facts included therein were drawn from the record and therefore neither represented nor were to be construed as conclusive findings of fact. *Id.* at 216 n.1.

2. *Id.* at 216. Dr. Boyce recommended lumbar surgery, specifically a decompression laminectomy and fusion at the L3-L4 and L4-L5 vertebrae. Banks consented to the procedure. *Id.*

3. *Id.* While Dr. Boyce's operative report indicated that he performed the procedure as intended, he realized thereafter that he had mistakenly performed surgery at the L2-L3 rather than the L4-L5 vertebrae.

4. *Id.*

5. *Id.*

6. All references herein to the "Elks Lodge" refer to the group of three defendants that collectively operated the lodge in which the plaintiff was injured: Elks Club Pride of Tennessee 1102, Pride of Tennessee Lodge of Elks No. 1102 Improved Benevolent, and Elks Lodge 1102 Pride of Tennessee. *See id.* at 216 & n.2. The third defendant, Elks Lodge 1102 Pride of Tennessee, owned the facility. *Id.* at 216 n.2.

7. *Id.* at 216.

8. *Id.*

consolidated.[9] In May 2008, the Elks Lodge filed a motion to amend its answer to assert the affirmative defense of comparative fault against Cumberland Manor, and shortly thereafter, Dr. Boyce did likewise, adopting the language of the Elks Lodge's motion.[10] Both defendants' motions alleged that the defendants "had learned during the discovery process that Cumberland Manor's improper care and treatment had contributed to Ms. Banks's staphylococcus infection," that the infection "had aggravated Ms. Banks's injuries and damages and that Ms. Banks was seeking to hold [the defendants] responsible for these additional injuries and damages."[11] The defendants also sought to reserve "the right to amend their comparative fault defense to allege fault of others throughout the course of discovery and trial."[12] The plaintiff opposed the defendants' motions.[13]

In August 2008, the trial court denied both motions, reasoning that the proposed amendments would be "futile" in light of precedent holding that tortfeasors cannot reduce their liability by alleging the subsequent negligence of medical providers.[14] The trial court did, however, suggest that the defendants should pursue an interlocutory appeal of its ruling.[15] The defendants did so, but the Tennessee Court of Appeals denied their application without comment.[16] In December 2008, the Tennessee Supreme Court granted the defendants permission to appeal, and immediately thereafter, the plaintiff amended her complaint to name Cumberland Manor as a defendant.[17]

The Tennessee Supreme Court, *held*, reversed and remanded: the trial court erred in not permitting Elks Lodge and Dr. Boyce to amend their answers to assert a comparative fault defense against Cumberland Manor.[18] Parties are liable for both the physical injuries they cause that necessitate medical treatment and subsequent injuries sustained during that treatment; however, liability for the subsequent injuries will be apportioned between the original tortfeasor and the subsequent healthcare providers according to

---

9. *Id.* at 216–17.

10. *Id.* at 217.

11. *Id.*

12. *Id.* (internal quotation marks omitted).

13. *Id.* The plaintiff argued that the defendants' efforts to attribute fault to Cumberland Manor for the latter's negligent medical treatment, which the defendants' negligence necessitated was inappropriate. *Id.*

14. *Id.* (discussing Transports, Inc. v. Perry, 414 S.W.2d 1 (Tenn. 1967); Atkinson v. Hemphill, No. 01-A-01-9311-CV00509, 1994 WL 456349 (Tenn. Ct. App. Apr. 24, 1994)).

15. *Id.*

16. *Id.*

17. *Id.* at 217–18. The plaintiff did this in order to avoid a potentially adverse statute of limitations impact should the Tennessee Supreme Court reverse the trial court's decision and permit the defendants to amend their answers to assert Cumberland Manor's comparative fault. *Id.*

18. *Id.* at 228.

fault.[19] *Banks v. Elks Club Pride of Tennessee 1102*, 301 S.W.3d 214 (Tenn. 2010).

II. ISSUE BEFORE THE COURT

Tennessee courts have long recognized the traditional "original tortfeasor rule."[20] The rule stipulates that the original tortfeasor's negligence, which initially injures the plaintiff, also proximately causes any injuries arising from subsequent medical treatment, making the original tortfeasor jointly and severally liable for both sets of injuries.[21] However, the continuing viability of the original tortfeasor rule became unclear when the Tennessee Supreme Court adopted a modified comparative fault system and, as a corollary, declared the joint and several liability doctrine "obsolete," except perhaps in "certain limited circumstances."[22]

Subsequent decisions addressed the extent to which comparative fault abolished joint and several liability but did not define the precise scope of the limited circumstances in which the doctrine remained viable.[23] Thus, the original tortfeasor rule's continuing viability likewise remained unclear.[24] In the *Banks* opinion, the supreme court first addressed this issue in the context of determining (1) whether the defendants, Elks Lodge and Dr. Boyce, could be held liable for medical providers' subsequent negligence in treating the injuries, and (2) whether each defendant could be held jointly and severally liable with subsequent negligent actors.[25]

III. DEVELOPMENT OF THE ORIGINAL TORTFEASOR RULE AND THE POST-*MCINTYRE* DOCTRINE OF JOINT AND SEVERAL LIABILITY

A. The Traditional Original Tortfeasor Rule

The traditional original tortfeasor rule embodies two common law principles.[26] The first principle is that, where a person "is injured by the negligence of another, and these injuries are aggravated by medical

---

19. *Id.*

20. *See id.* at 221–22.

21. *See* Transports, Inc. v. Perry, 414 S.W.2d 1, 4 (Tenn. 1967); Revell v. McCaughan, 39 S.W.2d 269, 271 (Tenn. 1931).

22. Banks, 301 S.W.3d at 218 (discussing McIntyre v. Balentine, 833 S.W.2d 52, 58, 60 (Tenn. 1992)).

23. *See, e.g.*, Owens v. Truckstops of Am., 915 S.W.2d 420, 430 (Tenn. 1996); Volz v. Ledes, 895 S.W.2d 677, 680 (Tenn. 1995); Bervoets v. Harde Ralls Pontiac-Olds, Inc., 891 S.W.2d 905, 907 (Tenn. 1994).

24. *See Banks*, 301 S.W.3d at 220–21.

25. *See id.* at 221.

26. *See* J.D. LEE & BARRY A. LINDAHL, MODERN TORT LAW: LIABILITY AND LITIGATION § 6:3 (2d ed. 2005).

treatment (either prudent or negligent), the negligence of the wrongdoer causing the original injury is regarded as the proximate cause of the damage subsequently flowing from the medical treatment."[27] The second principle is that the original tortfeasor is jointly and severally liable for the plaintiff's entire harm, including both the original injury and any subsequent injuries sustained during any necessary treatment.[28]

Designed to ensure that injured parties receive full compensation, the first principle eliminates the requirement that, to recover from an original tortfeasor for subsequent injuries, the plaintiff must prove that the original tortfeasor's conduct proximately caused the subsequent injuries.[29] In negligence claims, the proximate cause element requires that the harm be a "natural and probable consequence" of the defendant's conduct.[30] In Tennessee, the foreseeability of the harm that occurred is paramount in making this determination; even granting causation in fact, liability does not arise from an unforeseeable harm.[31] In most circumstances, the intentional intervention of a third party breaks the chain of causation required to hold an original tortfeasor liable: even if the original tort could have reasonably resulted in such an intervention, the original tortfeasor generally has no reason to expect that the intervening party will fail to exercise reasonable care.[32] The first principle precludes any such argument by automatically establishing the original injury as the proximate cause of additional injuries stemming from subsequent medical negligence.[33]

Tennessee courts have recognized and applied some form of the first principle for over one hundred years.[34] In 1900, the supreme court supported its holding that an injured plaintiff "is bound to exercise only reasonable and ordinary care in the employment of a physician" by citing cases from other jurisdictions in which defendants were held "responsible for full damages, although the physician so employed may have enhanced the damages by mistaken or unskillful treatment."[35] Since 1931, the first principle has changed very little, merely expanding in scope to apply where the subsequent medical treatment is "either prudent or negligent" and where

---

27. Transports, Inc. v. Perry, 414 S.W.2d 1, 4 (Tenn. 1967).

28. *See* LEE & LINDAHL, *supra* note 26, at § 6:3 n.1 (citing Lewis *ex rel.* Lewis v. Samson, 35 P.3d 972 (N.M. 2001)).

29. *See* Naifeh v. Valley Forge Life Ins. Co., 204 S.W.3d 758, 771 (Tenn. 2006) (outlining the elements of negligence).

30. Moody v. Gulf Ref. Co., 218 S.W. 817, 820 (Tenn. 1920) (citations omitted) (internal quotation marks omitted).

31. *See* Satterfield v. Breeding Insulation Co., 266 S.W.3d 347, 366 (Tenn. 2008) (citations omitted); Biscan v. Brown, 160 S.W.3d 462, 480 (Tenn. 2005) (citations omitted); Doe v. Linder Constr. Co., 845 S.W.2d 173, 178 (Tenn. 1992).

32. *See* 2 JACOB A. STEIN, STEIN ON PERSONAL INJURY DAMAGES § 11:7 (Gerald W. Boston ed., 3d ed. 1997).

33. *See id.*

34. *See* Ark. River Packet Co. v. Hobbs, 58 S.W. 278, 282 (Tenn. 1900).

35. *Id.*

non-physicians provide such treatment.[36] While initially applied only to disputes arising from workplace injuries, the first principle now "applies to the general field of tort law."[37]

The second principle of the original tortfeasor rule, which permits the plaintiff to obtain full recovery from the original tortfeasor, likewise reflects the goal of ensuring that injured parties receive full compensation.[38] This provision allows the plaintiff to be made whole merely by proving the original tortfeasor's negligence, whereas full recovery would otherwise require a more complex medical malpractice action.[39] Unlike the first principle, Tennessee precedent does not explicitly support the second principle.[40] While decisions involving the first principle have often applied the joint and several liability doctrine,[41] Tennessee courts have not considered the validity of such application since, prior to Tennessee's adoption of comparative fault, the doctrine was generally applicable: plaintiffs could fully recover from any properly joined defendant against whom they could establish liability.[42] Therefore, an original tortfeasor could avoid joint and several liability for the full extent of a plaintiff's injuries only by proving that the plaintiff's original injury did not necessitate the medical treatment during which the plaintiff sustained additional injuries.[43]

---

36. *Compare* Revell v. McCaughan, 39 S.W.2d 269, 271 (Tenn. 1931) ("[I]f one is injured by the negligence of another and these injuries are aggravated by the negligence of a physician, the negligence of the wrongdoer in causing the original injury is regarded as the proximate cause of the damage flowing from the subsequent negligent treatment by the physician."), *with* Transports, Inc. v. Perry, 414 S.W.2d 1, 4 (Tenn. 1967) ("[I]f one is injured by the negligence of another, and these injuries are aggravated by medical treatment (either prudent or negligent), the negligence of the wrongdoer causing the original injury is regarded as the proximate cause of the damage subsequently flowing from the medical treatment." (citing *Revell*, 39 S.W.2d 269)), *quoted in* Banks v. Elks Club Pride of Tenn. 1102, 301 S.W.3d 214, 217 n.3 (Tenn. 2010), *and* Estate of Jenkins, No. M2003-01561-COA-R3-CV, 2004 WL 2607531, at *5 (Tenn. Ct. App. Nov. 16, 2004) (applying the first principle where a nurse rendered subsequent medical treatment).

37. McAlister v. Methodist Hosp. of Memphis, 550 S.W.2d 240, 242 (Tenn. 1977) (citing *Transports*, 414 S.W.2d 1).

38. *See* Edwards v. Sisler, 691 N.E.2d 1252, 1255 (Ind. Ct. App. 1998) (citing Holden v. Balko, 949 F. Supp. 704, 711–12 (S.D. Ind. 1996)).

39. *See* Atkinson v. Hemphill, No. 01-A-01-9311-CV00509, 1994 WL 456349, at *2–3 (Tenn. Ct. App. Aug. 24, 1994).

40. *See, e.g.*, *Transports*, 414 S.W.2d 1; *Revell*, 39 S.W.2d 269.

41. *See, e.g.*, *Revell*, 39 S.W.2d at 270.

42. *See* TENN. R. CIV. P. 20.01 (permitting joinder of defendants where the plaintiff's right to relief from each stems from the "same transaction, occurrence, or series of transactions or occurrences and if any question of law or fact common to all defendants will arise in the action").

43. *See Transports*, 414 S.W.2d at 4.

B. The Scope of Joint and Several Liability Under the Comparative Fault Regime

In the 1992 case of *McIntyre v. Balentine,* the Tennessee Supreme Court abandoned the contributory negligence doctrine in favor of a modified comparative fault system: plaintiffs' damages would be reduced in proportion to the percentage of their total negligence, but plaintiffs could only recover if their negligence was less than the defendants'.[44] In its holding, the court recognized that the adoption of comparative fault would "affect[] numerous legal principles surrounding tort litigation" but concluded that "[f]or the most part, harmonizing these principles with comparative fault . . . should await an appropriate controversy."[45] The court did, however, provide some "guidance" to courts in implementing the new system.[46] This guidance included declaring that the *McIntyre* holding rendered the joint and several liability doctrine "obsolete":[47]

> Our adoption of comparative fault is due largely to considerations of fairness: the contributory negligence doctrine unjustly allowed the entire loss to be borne by a negligent plaintiff, notwithstanding that the plaintiff's fault was minor in comparison to [the] defendant's. Having thus adopted a rule more closely linking liability and fault, it would be inconsistent to simultaneously retain a rule, joint and several liability, which may fortuitously impose a degree of liability that is out of all proportion to fault.[48]

The court did, however, leave open the possibility that joint and several liability might still apply in "certain limited circumstances."[49]

The *McIntyre* court's rationale for adopting a modified comparative fault rule, which subsequent courts reiterated as the comparative fault regime's overarching goal, was to strike the proper balance between "the plaintiff's interest in being made whole" and "the defendant's interest in paying only that percentage of damages for which that particular defendant

---

44. McIntyre v. Balentine, 833 S.W.2d 52, 57 (Tenn. 1992).

45. *Id.* at 57, 60.

46. *Id.* at 57–58.

47. *Id.* at 58.

48. *Id.* This declaration was later characterized as dictum. *See* Owens v. Truckstops of Am., 915 S.W.2d 420, 435 (Tenn. 1996) (Drowota, J., dissenting); Volz v. Ledes, 895 S.W.2d 677, 680 (Tenn. 1995); Owens v. Truckstops of Am., No. 01A-01-9305-CV-00208, 1994 WL 115878, at *11–12 (Tenn. Ct. App. Apr. 6, 1994) (classifying it as judicial dictum "because it was made with consideration and purpose for the benefit and guidance of the bench and bar"). However, subsequent holdings repeated it. *See Volz*, 895 S.W.2d at 680; Bervoets v. Harde Ralls Pontiac-Olds, Inc., 891 S.W.2d 905, 907 (Tenn. 1994); *see also* Camper v. Minor, 915 S.W.2d 437, 448 n.5 (Tenn. 1996) (discussing the weight of the *Volz* and *Bervoets* holdings).

49. *McIntyre*, 833 S.W.2d at 60.

is responsible."[50] As the *McIntyre* court suggested,[51] this balance has generally been found by linking liability to fault.[52]

One year after the *McIntyre* decision, the Tennessee General Assembly remedied a potential problem accompanying the comparative fault regime by legislating that, where a defendant alleges a nonparty's comparative fault and where the statute of limitations would otherwise bar the plaintiff's cause of action against the nonparty, the plaintiff may either amend its complaint to add the nonparty as a defendant or institute a separate action against the nonparty within ninety days.[53]

In the 1994 case of *Bervoets v. Harde Ralls Pontiac-Olds, Inc.*, the supreme court first considered the competing interests of multiple tortfeasors and affirmed that *McIntyre* abolished joint and several liability "to the extent that it allows a plaintiff to sue and obtain a full recovery against any one or more of several parties against whom liability could be established."[54] Specifically, the *Bervoets* court reasoned that the statutory remedy of contribution, permitting defendants who have paid more than their fair share of a judgment to recover from codefendants, would rarely be necessary under the comparative fault system.[55] The court concluded that the remedy nevertheless remained viable in the "appropriate case[s]" and that, in such cases, it would order contribution according to comparative fault principles.[56]

In the 1995 case of *Volz v. Ledes*, the supreme court reaffirmed that *McIntyre* rendered the joint and several liability doctrine obsolete.[57] Considering circumstances in which a decedent and two successive physicians were negligent, the court held that the second physician's negligence should be assessed as a proximate cause in the wrongful death action rather than in a (therein proposed) separate "loss of chance" cause of action.[58] In light of the *McIntyre* court's goal of linking liability to fault, the court rejected a proposed exception permitting a defendant's liability to be enhanced beyond that defendant's percentage of fault because such a rule would allow the "happenstance of the financial wherewithal of other defendants" to determine a defendant's liability.[59] Later that year in

50. Brown v. Wal-Mart Disc. Cities, 12 S.W.3d 785, 787 (Tenn. 2000).

51. *McIntyre*, 833 S.W.2d at 58.

52. *See* Biscan v. Brown, 160 S.W.3d 462, 474 (Tenn. 2005); Ali v. Fisher, 145 S.W.3d 557, 563–64 (Tenn. 2004); Carroll v. Whitney, 29 S.W.3d 14, 16–17 (Tenn. 2000).

53. TENN. CODE ANN. § 20-1-119 (2009).

54. Bervoets v. Harde Ralls Pontiac-Olds, Inc., 891 S.W.2d 905, 907 (Tenn. 1994).

55. *Id.* (quoting *McIntyre*, 833 S.W.2d at 58).

56. *Id.* at 907, 908 (discussing TENN. CODE ANN. §§ 29-11-101 to -106 (1980)).

57. Volz v. Ledes, 895 S.W.2d 677, 680 (Tenn. 1995).

58. *Id.* at 679–80. This was later clarified as one of a limited number of circumstances in which a healthcare provider's liability may be reduced or avoided by asserting the patient's comparative fault. Mercer v. Vanderbilt Univ., Inc., 134 S.W.3d 121, 129 (Tenn. 2004).

59. *Volz*, 895 S.W.2d at 680.

Whitehead v. Toyota Motor Corp., the Tennessee Supreme Court again confirmed the joint and several liability doctrine's obsolescence.[60] Holding that comparative fault principles applied to products liability actions based on strict liability, the court permitted a defendant manufacturer to assert a comparative fault defense against a plaintiff where an alleged defect enhanced the plaintiff's injuries but did not cause or contribute to the underlying accident.[61]

The majority opinion in *Owens v. Truckstops of America*, the first post-*McIntyre* case permitting joint and several liability, represented a significant departure from the supreme court's prior indications regarding how Tennessee courts should interpret *McIntyre*.[62] As in *Whitehead*, the court considered the applicability of comparative fault to strict liability actions, but while the issue in *Whitehead* pertained to the allocation of fault between the plaintiff and defendant, the *Owens* court considered whether multiple defendants in a product's chain of distribution were jointly and severally liable.[63] Citing *McIntyre* and *Bervoets*, the *Owens* court declared that the concepts of fairness and efficiency are the basis of comparative fault.[64]

Examining post-*McIntyre* decisions, the court reached two broad conclusions.[65] The court first concluded that these earlier cases involved circumstances in which applying comparative fault did not "work[] an unfairness upon any party."[66] The court distinguished the present case as one in which fairness considerations necessitated departing from comparative fault principles in favor of joint and several liability.[67] Though invoking the *McIntyre* court's fairness considerations and statement that the joint and several liability doctrine might apply in certain limited circumstances, the court did not expressly deem its holding an exception to the doctrine's general obsolescence.[68] Rather, the court reviewed the facts of prior cases interpreting the *McIntyre* decision and therefrom reached a second conclusion regarding the general precedential scope of post-*McIntyre* cases: "[W]here the *separate, independent negligent acts* of more than one tortfeasor combine to cause a *single, indivisible injury*, each

60. Whitehead v. Toyota Motor Corp., 897 S.W.2d 684, 686 (Tenn. 1995).

61. *Id.* at 693–94.

62. *See* Owens v. Truckstops of Am., 915 S.W.2d 420, 425–26 (Tenn. 1996).

63. *See id.* at 424.

64. *Id.* at 424, 430.

65. *Id.* at 425–26, 430.

66. *Id.* at 425–26.

67. *Id.* at 426. This conclusion was based on the fact that, had the court allocated fault to the two nonparties to the plaintiff's case, the plaintiff would not have been able to recover from these nonparties. *Id.* Tennessee Code Annotated section 20-1-119 did not apply because this statute was not enacted until after the statute of limitations had barred the plaintiff's claims. *Id.* at 427.

68. *Id.* at 428, 430 (quoting *McIntyre*, 833 S.W.2d at 60).

tortfeasor will be liable only for that proportion of the damages attributable to its fault" (the *Owens* rule).[69]

Because post-*McIntyre* holdings had only affirmed the obsolescence of joint and several liability in cases involving these limited circumstances, the *Owens* court clarified that it had not "disapproved of the doctrine of joint and several liability *in a general sense*" but merely in this "*particular sense*."[70] Dissenting in part, Justice Drowota, the author of *McIntyre*'s majority opinion claimed that this decision "resurrect[ed]" the obsolete joint and several liability doctrine.[71] The majority responded that it did not need to resurrect the doctrine because the doctrine continued to be an "integral part of the law, except where specifically abrogated."[72]

Subsequent supreme court decisions accorded with the *Owens* rule's specific corollary that the joint and several liability doctrine no longer applies to circumstances in which multiple tortfeasors' separate, independent acts of negligence combine to cause a single, indivisible injury (the *Owens* corollary).[73] Semantically, the court has generally reaffirmed its past declarations—which the dissent in *Owens* cited—that the joint and several liability doctrine is obsolete, but the circumstances in which the court has cited these declarations as grounds for disallowing the doctrine's application generally fall within the scope of the *Owens* corollary.[74] Conversely, subsequent exceptions to the general obsolescence of joint and several liability have involved circumstances that fall outside of the *Owens* corollary's scope: while the court did not always describe these exceptions in the corollary's terms, each might reasonably be viewed as *not* involving separate, independent acts of negligence.[75]

---

69. *Id.* at 430 (emphasis added).

70. *See id.* at 431 n.13.

71. *Id.* at 437 (Drowota, J., dissenting).

72. *Id.* at 431 n.13.

73. *Id.* at 430; *see* Sherer v. Linginfelter, 29 S.W.3d 451, 455 (Tenn. 2000); Resolution Trust Corp. v. Block, 924 S.W.2d 354, 355 (Tenn. 1996).

74. *See, e.g.*, Ali v. Fisher, 145 S.W.3d 557, 561, 565–66 (Tenn. 2004) (referring to the abandonment of joint and several liability as a "settled principle[]" and upholding the allocation of fault for a car accident injury between two defendants, the negligent driver and the owner who negligently entrusted the vehicle); Carroll v. Whitney, 29 S.W.3d 14, 21 (Tenn. 2000) (refusing to permit an "implicit revival of joint and several liability" and upholding the allocation of fault for medical malpractice between negligent resident physicians who acted independently); *Sherer*, 29 S.W.3d at 455 (reiterating the *Owens* rule and joint and several liability's obsolescence and requiring allocation of fault for a car accident injury between two defendants, the negligent driver and the vehicle's manufacturer whose liability enhanced the injury). *Cf. Resolution Trust*, 924 S.W.2d at 355 (quoting *Owens*, 915 S.W.2d at 431 n.13) (reiterating that joint and several liability remains an "integral part of the law" and permitting it to be applied to corporate employees who collectively breached a common duty).

75. *See* Limbaugh v. Coffee Med. Ctr., 59 S.W.3d 73, 87 (Tenn. 2001) (holding that a tortfeasor whose duty it is to protect others from third persons' foreseeable, intentional acts

Though not affecting the *Owens* corollary, the court's holding in *Mercer v. Vanderbilt University, Inc.* merits discussion as an apparent exception to the principal *Owens* rule.[76] Therein, the supreme court held that "a patient's negligent conduct that occurs prior to a health care provider's negligent treatment and provides only the occasion for the healthcare provider's subsequent negligence may not be compared to the negligence of the health care provider."[77] The circumstances in *Mercer* indisputably involved the separate, independent acts of a plaintiff, who negligently caused his car accident, and a hospital, which provided negligent treatment after the accident.[78] In considering the applicability of the *Owens* rule, the court reassessed its holding in *Gray v. Ford Motor Co.*, an earlier case involving analogous circumstances.[79] Holding that such circumstances are governed by the *Owens* rule, the *Gray* court ordered the allocation of fault between the plaintiff and physician, reasoning that the comparative fault principles underlying the *Owens* rule do not change where the plaintiff's negligence is a contributing proximate cause of the injury.[80] In overruling the *Gray* decision, the *Mercer* court did not directly address the *Owens* rule, but its holding appeared to be a clear exception thereto based on policy considerations.[81] The court reasoned that, regardless of whether the plaintiff's injuries were indivisible or separate,[82] "it would be unfair to allow a defendant doctor to complain about the patient's negligence because this negligence caused the very condition the doctor undertook to treat."[83] Specifically, the court reasoned that the *Gray* precedent, which might exculpate otherwise liable medical providers based

---

is jointly and severally liable with the third persons for injuries that the third persons' intentional acts caused); Browder v. Morris, 975 S.W.2d 308, 311–12 (Tenn. 1998) (citing Camper v. Minor, 915 S.W.2d 437, 447–48 (Tenn. 1996)) (holding that, to the extent that vicarious liability can be considered species of joint and several liability, *McIntyre* did not affect vicarious liability doctrines, such as the family purpose doctrine, respondeat superior, and others applicable in "similar circumstance[s] where liability is vicarious due to an agency-type relationship"); Gen. Elec. Co. v. Process Control Co., 969 S.W.2d 914, 916–17 (Tenn. 1998) (holding that defendants whose concerted action causes a single, indivisible injury are jointly and severally liable); *Resolution Trust*, 924 S.W.2d at 355, 357 (holding that defendants who breached a common duty are jointly and severally liable); *Owens*, 915 S.W.2d at 433 (holding that defendants in a product's distribution chain in a products liability action are jointly and severally liable).

76. *See* Mercer v. Vanderbilt Univ., Inc., 134 S.W.3d 121, 130 (Tenn. 2004).

77. *Id.*

78. *See id.* at 125–27.

79. *Id.* at 127–28 (discussing Gray v. Ford Motor Co., 914 S.W.2d 464 (Tenn. 1996)).

80. *See Gray*, 914 S.W.2d at 467 (citing Whitehead v. Toyota Motor Corp., 897 S.W.2d 684 (Tenn. 1995)).

81. *See Mercer*, 134 S.W.3d at 125–32.

82. *Id.* at 128.

83. *Id.* at 129.

on the cause of the initial injury, was anomalous with the otherwise uniform standard of care to which such providers are held.[84]

C. The Original Tortfeasor Rule Under the Comparative Fault Regime

During the eighteen-year span between the *McIntyre* and *Banks* decisions, each section of the Tennessee Court of Appeals addressed the original tortfeasor rule's continuing viability in light of the adoption of comparative fault.[85] In *Atkinson v. Hemphill*, the court of appeals held that the traditional rule was unaffected but did not ratify each constituent principle[86] in isolation from the other; instead, it addressed the second principle as a corollary of the first.[87] The *Atkinson* court reasoned that the traditional rule's application of joint and several liability was based on the rationale that "the tortfeasor whose negligence caused the injured party to require medical attention should bear all the foreseeable risks resulting from the injury."[88] Thus, the original tortfeasor "shoulder[ed] the difficult task of determining whether the medical provider . . . committed professional malpractice" but retained the right to recover from the negligent medical provider by either impleading the provider in the original suit or suing the provider thereafter under a theory of subrogation.[89]

Considering the *McIntyre* decision's effect on the original tortfeasor rule as an indivisible unit, the court of appeals declared, "[T]o allow a tortfeasor to reduce his damages by alleging [as an affirmative defense] the subsequent negligence of a medical provider would for all practical purposes abolish the . . . rule."[90] The court reasoned that abolition would "penalize injured parties in several inequitable ways."[91] It found that allowing defendants to reduce their liability by convincing the trier of fact to apportion fault to subsequently negligent treatment providers after the statute of limitations had barred the plaintiffs' claims against such providers would compel plaintiffs to name all subsequent medical providers as defendants—whether or not they suspected malpractice—so as not to risk forfeiting a full recovery.[92] Therefore, the court of appeals reasoned,

---

84. *Id.* at 129–30 (quoting Harvey *ex rel.* Harvey v. Mid-Coast Hosp., 36 F. Supp. 2d 32, 38 (D. Me. 1999)).

85. *See* Jackson v. Hamilton, No. W2000-01992-COA-R3-CV, 2003 WL 22718386, at *5–6 (Tenn. Ct. App. Nov. 4, 2003); Troy v. Herndon, No. 03A01-9707-CV-00271, 1998 WL 820698, at *1–2 (Tenn. Ct. App. Nov. 24, 1998); Atkinson v. Hemphill, No. 01-A-01-9311-CV00509, 1994 WL 456349, at *2 (Tenn. Ct. App. Aug. 24, 1994).

86. *See Atkinson*, 1994 WL 456349, at *2.

87. *See id.*

88. *Id.*

89. *Id.*

90. *Id.*

91. *Id.* at *3.

92. *Id.* The court of appeals did not consider the effect of TENN. CODE ANN. § 20-1-119 (2009). *See supra* text accompanying notes 53–59.

abolishing the rule would effectively shift the burden of proof to the plaintiff.[93] The *Atkinson* court concluded that the supreme court did not intend to penalize injured plaintiffs in this manner and thus held that the *McIntyre* decision did not affect the common law original tortfeasor rule.[94]

In *Troy v. Herndon* and *Jackson v. Hamilton*, the eastern and western sections of the court of appeals addressed the original tortfeasor rule's viability.[95] Both courts drew heavily from the *Atkinson* opinion in holding that *McIntyre* did not abolish the traditional original tortfeasor rule.[96] While no defendant in either the *Atkinson* or the *Troy* case sought permission to appeal to the Tennessee Supreme Court, a defendant in *Jackson* sought such permission.[97] The supreme court, while declining to review the decision, designated the opinion "Not for Citation," thereby signifying it had no precedential value.[98] The court of appeals later declined to address the original tortfeasor rule in *Estate of Jenkins*, inferring from this designation that the supreme court had "serious questions" about the rule's continuing viability.[99]

IV. AN ORIGINAL TORTFEASOR'S LIABILITY FOR ENHANCED DAMAGES CAUSED BY SUBSEQUENT MEDICAL NEGLIGENCE MAY BE REDUCED BY ALLOCATING FAULT TO THE NEGLIGENT HEALTHCARE PROVIDER

In *Banks*, the Tennessee Supreme Court unanimously held that the defendants, Elks Lodge and Dr. Boyce, could assert a comparative fault defense against subsequent healthcare providers alleged to have negligently provided such care.[100] The court based its decision on these grounds: (1) the *McIntyre* decision did not affect the original tortfeasor rule's first principle; (2) a party could be labeled both an original and a successive tortfeasor; (3) Tennessee's adoption of a comparative fault regime disapproved the second

---

93. *Id.*

94. *Id.* at *2. Instead, the court interpreted the *McIntyre* decision as simply "eliminat[ing] the harsh rule that a plaintiff's contributory negligence completely bars his recovery, and . . . establish[ing] a system of comparative fault that more equitably allocates liability between negligent actors." *Id.* at *3.

95. *See* Jackson v. Hamilton, No. W2000-01992-COA-R3-CV, 2003 WL 22718386, at *5–6 (Tenn. Ct. App. Nov. 4, 2003); Troy v. Herndon, No. 03A01-9707-CV-00271, 1998 WL 820698, at *1–2 (Tenn. Ct. App. Nov. 24, 1998).

96. *See Jackson*, 2003 WL 22718386, at *5–6; *Troy*, 1998 WL 820698, at *1–3.

97. *See* Estate of Jenkins, No. M2003-01561-COA-R3-CV, 2004 WL 2607531, at *5 (Tenn. Ct. App. Nov. 16, 2004).

98. Jackson v. Hamilton, No. W2000-01992-SC-R11-CV 2003 WL 22718386, at *5–6 (Tenn. Ct. App. Nov. 4, 2003), perm. app. denied, designated not for citation (Tenn. May 10, 2004).

99. Estate of Jenkins, 2004 WL 2607531, at *5.

100. Banks v. Elks Club Pride of Tenn. 1102, 301 S.W.3d 214, 228 (Tenn. 2010).

principle's imposition of joint and several liability; and (4) public policy did not support retaining joint and several liability in this circumstance.[101]

Justice Koch, writing for the majority, began by outlining the context of this decision, part of the "sea-change" in Tennessee's tort law that the *McIntyre* court initiated by adopting comparative fault.[102] The *Banks* decision represented the most recent development in a line of cases interpreting the *McIntyre* court's declaration that comparative fault rendered the joint and several liability doctrine "obsolete" but that the doctrine might be retained "in certain limited circumstances."[103] However, this was the first occasion where the supreme court addressed the original tortfeasor rule's continuing viability in light of *McIntyre*.[104]

The court began its analysis by examining the appellate court's decisions in *Atkinson*, *Troy*, and *Jackson*.[105] First, the supreme court affirmed the appellate court's holdings regarding the first principle.[106] The court reasoned that tortfeasors are generally liable not only for injuries caused by their own negligence but also for any subsequent injuries caused by another tortfeasor's negligent conduct when the conduct giving rise to the subsequent injury was a "foreseeable or natural consequence of the original tortfeasor's conduct."[107] Most commonly invoked in cases of negligence in subsequent treatment of tortiously caused injuries,[108] this principle has been recognized in Tennessee for over a century.[109] The court clarified that, because this principle establishes *when* defendants are liable, the *McIntyre* decision's restrictions on the *extent of harm* for which defendants are liable did not affect it.[110] Thus, the court held that Tennessee's rule embodying the first principle remains that:

> [A]n actor whose tortious conduct causes physical harm to another is liable for any enhanced harm the other suffers due to the efforts of third persons to render aid reasonably required by the other's injury, as long as the enhanced harm arises from a risk that inheres in the effort to render aid.[111]

---

101. *Id.* at 226–28.

102. *Id.* at 218.

103. *Id.* (quoting McIntyre v. Balentine, 833 S.W.2d 52, 58, 60 (Tenn. 1992)).

104. *Id.* at 220.

105. *Id.* at 221.

106. *Id.* at 221–22. However, the court found that the appellate court nevertheless "failed to differentiate between the two principles." *Id.* at 221.

107. *Id.* at 222.

108. *Id.*

109. *See supra* notes 34–37 and accompanying text.

110. *Banks*, 301 S.W.3d at 223 (citing RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL HARM § 35 cmt. d (Proposed Final Draft No. 1 2005)).

111. *Id.*; *see* RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL HARM § 35 (Proposed Final Draft No. 1 2005).

Next, the supreme court disapproved the appeals court's holdings regarding the second principle, which determines the apportionment of liability among multiple tortfeasors.[112] The court declared that the *Atkinson*, *Troy*, and *Jackson* courts had erred in failing to differentiate between and consider independently the original tortfeasor rule's two principles.[113] Based on this erroneous premise, the appeals court had, in each case, faced the decision of either rejecting the rule as a whole—including the well-established first principle—or accepting the rule as a whole, including the second principle, which contravenes the *Owens* court's abolishment of joint and several liability in these circumstances.[114] The supreme court disapproved the appeals court's decisions in *Atkinson, Troy*, and *Jackson* because it determined that this "all-or-nothing" approach was unnecessary.[115] While pure several liability "limits the liability of each defendant liable for the same harm to that defendant's comparative share of the harm," it does not "provide rules about when defendants are liable for harm that they caused."[116] Therefore, the court concluded that its decision to retain the first principle was consistent with its decision that the *Owens* corollary, which abolishes joint and several liability where "separate, independent negligent acts combine to cause a single, indivisible injury," overruled the second principle.[117] In support of this holding, the court cited other states' decisions drawing the same conclusion: "comparative fault does not prevent the continuing imposition of liability on an original tortfeasor for subsequent negligent medical care for the injuries caused by the original tortfeasor."[118]

The court also addressed the assertions of the plaintiff and the Tennessee Association for Justice (as amicus curiae) that public policy dictated retaining joint and several liability in this instance.[119] First, the parties asserted that "joint and several liability [was] appropriate [in such circumstances] because the original defendant [was] the proximate cause of the entire injury."[120] The court rejected this argument, explaining that "the original tortfeasor's conduct [was] not the *sole* proximate cause of the plaintiff's indivisible injury," but rather the "independent tortious conduct of the original tortfeasor and [of the successive tortfeasor were] both

112. *Id.* at 221–23.

113. *Id.* at 221.

114. *See* Atkinson v. Hemphill, No. 01-A-01-9311-CV00509, 1994 WL 456349, at *2–3 (Tenn. Ct. App. Aug. 24, 1994).

115. *Banks*, 301 S.W.3d at 221.

116. *Id.* at 223 (quoting RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL HARM § 35 cmt. d (Proposed Final Draft No. 1 2005) (internal citation omitted)).

117. *Id.*

118. *Id.* at 224.

119. *Id.*

120. *Id.*

proximate causes of the injury."[121] Although the court had permitted joint and several liability in circumstances where multiple parties' conduct created multiple proximate causes,[122] such circumstances did not involve separate, independent acts and thus fell outside the scope of the *Owens* corollary.[123]

Second, the plaintiff and amicus curiae asserted an argument echoing the *Atkinson* court's concerns: abolishing joint and several liability would "place plaintiffs in the difficult position of being forced to sue their treating physicians and, thereby, [would add] the complexity of a medical negligence claim to an otherwise straightforward ordinary negligence case."[124] The court found that these concerns were overstated and outlined the proper procedure for when a tortfeasor asserts the affirmative defense that a nonparty healthcare provider is at fault for the plaintiff's injuries.[125] The plaintiff's first option is to not name the healthcare provider as a defendant.[126] As the *Atkinson* court addressed,[127] the plaintiff would then risk a diminished recovery if the defendant could convince the trier of fact that the nonparty provider is partially or completely at fault.[128] However, the *Atkinson* court did not consider the plaintiff's second option: to amend its complaint to add the nonparty provider as a defendant pursuant to Tennessee Code Annotated section 20-1-119.[129] The court clarified that, contrary to the plaintiff and amicus curiae's assertions (and perhaps in response to the *Atkinson* court's general concerns),[130] the burden of proving medical negligence "does not shift entirely to the plaintiff" in such circumstances but "remains with the original defendant who asserted the affirmative defense of comparative fault."[131] Alluding to the traditional rule's rationale,[132] the court concluded that a plaintiff "is not required to shoulder the difficulty and expense of proving medical negligence unless,

---

121. *Id.* (emphasis added).

122. *See supra* note 75 and accompanying text.

123. *Banks*, 301 S.W.3d at 224.

124. *Id.* at 224–25; *see also* Atkinson v. Hemphill, No. 01-A-01-9311-CV00509, 1994 WL 456349, at *3 (Tenn. Ct. App. Aug. 24, 1994).

125. *Banks*, 301 S.W.3d at 225.

126. *Id.*

127. *Atkinson*, 1994 WL 456349, at *3.

128. *Banks*, 301 S.W.3d at 225.

129. *Id.*; *see also supra* note 53 and accompanying text.

130. *See supra* notes 91–92 and accompanying text. The *Atkinson* court expressed concerns about effectively shifting the burden of proving medical negligence to the plaintiff in the context of rejecting a holding compelling a plaintiff to preemptively sue subsequent healthcare providers. *Atkinson*, 1994 WL 456349, at *3. The court did not consider circumstances in which a defendant asserted a nonparty healthcare provider's comparative fault and a plaintiff subsequently brought suit against that provider; thus, the court did not express these concerns about shifting the burden of proof in this context.

131. *Banks*, 301 S.W.3d at 225.

132. *See Atkinson*, 1994 WL 456349, at *2.

for some reason, it chooses to do so."[133] Pursuant to this conclusion, the court held that, under these circumstances, a trial court cannot act on a motion for a directed verdict filed by the newly added defendant until "after the close of all proof."[134] The court found that:

> If the original defendant is unable to prove that the healthcare provider is liable, the plaintiff may still obtain a complete recovery from the original defendant . . . . If, however, the original defendant is successful in proving that the healthcare provider is liable, then the plaintiff may obtain a complete recovery apportioned between the original defendant and the [negligent] provider based on their fault.[135]

The court reasoned that this rule prevented prejudice to plaintiffs who elect to amend their complaints by eliminating any compulsion for them to prove subsequent medical negligence.[136] Therefore, the plaintiff's decision whether or not to amend would remain entirely in his control, and if the plaintiff did not elect to amend, he would not be forced to "try a case it was not prepared to try."[137]

The court then addressed Elks Lodge's argument that under *Mercer*, "a negligent actor should not be held responsible for the subsequent negligence of a healthcare provider" and that the doctrine of stare decisis precluded the original tortfeasor rule's application therein.[138] The court rejected this argument for two reasons.[139] First, Elks Lodge's argument overstated the scope of the *Mercer* holding.[140] According to the *Banks* court, the rationale underlying the *Mercer* decision was that injured patients found to be more than fifty percent at fault should not forfeit recovery entirely.[141] Because the issue before the *Banks* court was the allocation of fault between negligent defendants, this policy concern did not apply; applying comparative fault in this case would not penalize an injured plaintiff but would merely "enable the trier of fact to apportion fault

---

133. *Banks*, 301 S.W.3d at 225.

134. *Id.* at 225 n.14. Because the burden of proof remains with the defendant, a directed verdict at the close of the plaintiff's proof would be inappropriate except "when the original defendant states that it lacks sufficient evidence to send the issue of the new defendant's fault to the jury. If the new defendant's motion for directed verdict is granted, the jury cannot . . . allocate any portion of the fault to the now-dismissed defendant." *Id.*

135. *Id.* at 225.

136. *Id.*

137. *Id.*

138. *Id.* at 226 (internal quotation marks omitted) (discussing Mercer v. Vanderbilt Univ., Inc., 134 S.W.3d 121 (Tenn. 2004)).

139. *Id.*

140. *Id.*

141. *Id.* (citing *Mercer*, 134 S.W.3d at 129–30).

between the defendants whose conduct caused or contributed to the plaintiff's injuries."[142]

Second, the court held that Elks Lodge's interpretation of the *Mercer* holding was "contrary to the basic tenets of Tennessee tort law, more than one century of Tennessee common law precedents, and the general principles of liability."[143] The original tortfeasor rule's first principle is rooted in the common law tort of negligence.[144] Negligent parties are "liable for the natural and probable consequences of their conduct,"[145] the scope of which is determined primarily by the foreseeability of such an injury[146] and whether the conduct was a "substantial factor" in bringing about the injury.[147] Had the trier of fact found that Elks Lodge's negligence substantially caused foreseeable injuries to Ms. Banks, an interpretation of *Mercer* exculpating the Elks Lodge from liability would contravene the basic tenets of negligence.[148] Because neither argument by the Elks Lodge was persuasive, the court concluded that the *Mercer* holding did not dictate a contrary result and declined to extend its scope so as to effectively eliminate the original tortfeasor rule.[149]

Finally, the court addressed Dr. Boyce's argument that he could not be deemed both an original tortfeasor and a successive tortfeasor.[150] Based on this apparent paradox, Dr. Boyce argued that the original tortfeasor rule should not apply to him so as to deem his negligence the proximate cause of Ms. Banks's subsequent injuries flowing from her treatment at Cumberland Manor.[151] The court rejected this argument, adopting the approach that the

---

142. *Id.*

143. *Id.*

144. *Id.*

145. *Id.* (citing Doe v. Linder Constr. Co., 845 S.W.2d 173, 181 (Tenn. 1992)).

146. *See supra* note 31 and accompanying text.

147. *Banks*, 301 S.W.3d at 226 (citing Naifeh v. Valley Forge Life Ins. Co., 204 S.W.3d 758, 771 (Tenn. 2006)).

148. *See id.* For the Elks Lodge to be held liable for Ms. Banks's original injuries, the plaintiff would have had to prove that the Elks Lodge should have foreseen these original injuries. *See, e.g.*, *Doe*, 845 S.W.2d at 181. Had the plaintiff successfully proved this foreseeability, the Elks Lodge would have been liable for the subsequent injuries: its conduct created a foreseeable risk of harm and thus created the foreseeable risk that such harm would require medical treatment. *See* LEE & LINDAHL, *supra* note 26. Thus, the original tortious conduct proximately caused the subsequent injuries sustained during treatment as long as (1) the treatment administered was "reasonably required to cure the injuries" caused by the original tort, and (2) the subsequent injuries resulted from "a risk inherent" in this treatment. *Id.*

149. *Banks*, 301 S.W.3d at 227.

150. *Id.* at 227–28 (internal quotation marks omitted). Dr. Boyce made four other arguments, but the court concluded that these did not merit discussion because they, respectively, were "beyond the scope of the issue certified on the interlocutory appeal," had been waived, and were moot in light of the other holdings. *Id.* at 227 n.16.

151. *Id.* at 227.

Missouri Court of Appeals employed in *State ex rel. Blond v. Stubbs*.[152] In that case, the Missouri court held that, where a corporation caused an initial injury and three doctors successively caused additional injuries, each tortfeasor was liable for both the injuries caused by the tortfeasor's direct negligence and those sustained during the subsequent medical treatment that the tortfeasor's negligence necessitated.[153] Therefore, the first-in-time physician was a successive tortfeasor as to the initial injury and an original tortfeasor as to both the injuries he caused and the injuries the later-in-time physicians caused; the second-in-time physician was a successive tortfeasor as to the initial injury as well as the injuries the first-in-time physician caused and was an original tortfeasor as to both the injuries he caused and the injuries the third-in-time physician caused.[154] In light of this example, the Tennessee Supreme Court concluded that a party's "continuing liability under the original tortfeasor rule is not tied to anything magical about being the 'original' tortfeasor"; rather, it results from the party "being a proximate cause of an aggravated injury resulting from subsequent medical treatment."[155] Therefore, the court held that Dr. Boyce could simultaneously be an original tortfeasor as to any aggravation of Ms. Banks's injury that he caused as well as any subsequent aggravation that Cumberland Manor's negligence caused and a successive tortfeasor as to the initial injury that Elks Lodge caused.[156]

The court concluded by reaffirming the general validity of both the original tortfeasor rule's first principle and the *Owens* corollary.[157] Because circumstances within the first principle's scope are necessarily within the *Owens* corollary's scope, the *Owens* corollary overruled the second principle's imposition of joint and several liability under the rule.[158] Applying this reasoning to the issue at hand, the court held that "the trial court erred by refusing to permit the Elks Lodge defendants and Dr. Boyce to amend their answers to assert a comparative fault defense against Cumberland Manor."[159]

V. IMPLICATIONS OF BANKS V. ELKS CLUB PRIDE OF TENNESSEE 1102

The Tennessee Supreme Court's decision in *Banks* reflects the inherent difficulties that accompany a state's adoption of a comparative fault regime founded on a goal of fully proportional liability: such a regime necessitates

152. *Id.* at 227–28 (citing State *ex rel.* Blond v. Stubbs, 485 S.W.2d 152, 154 (Mo. Ct. App. 1972)).

153. *Blond*, 485 S.W.2d at 154.

154. *Banks*, 301 S.W.3d at 227 (quoting *Blond*, 485 S.W.2d at 154).

155. *Id.* at 227–28.

156. *Id.* at 228.

157. *Id.*

158. *Id.*

159. *Id.*

the widespread reassessment of the state's statutory and common law to extricate and overrule only those principles that are explicitly contrary to this goal.[160] As a direct result of the court's holding, defendants whose conduct necessitates subsequent medical treatment can now lessen their liability for injuries sustained during such treatment by asserting healthcare providers' comparative fault.[161] However, given the court's procedural clarification, statute of limitations concerns will no longer compel plaintiffs to sue these providers preemptively to ensure full recovery; the burden of proving medical malpractice will remain with defendants who assert the comparative fault defense.[162]

However, the court's opinion suggests that the scope of its holding will not be limited to these direct ramifications. Rather, it appears that the court intended to create a broad framework for the application of comparative fault in future cases.

A. Comparative Fault Principles Will Supplant Common Law Rules, but Only Where Explicitly Necessary to Ensure Proportional Liability

Regardless of the stated goal of Tennessee's comparative fault regime—to balance "the plaintiff's interest in being made whole with the defendant's interest in paying only those damages for which the defendant is responsible,"[163]—a state's adoption of comparative fault necessarily entails that the latter interest will take precedence.[164] Therefore, under Tennessee's comparative fault regime, the latter policy of proportional liability has continued to prevail where the two policies are irreconcilable.[165] However, the supreme court's stated goal of balancing interests and the manner in which it did so in *Banks* both suggest that the plaintiff's interest in full recovery remains relevant to the extent that it can be reconciled with the essential tenets of comparative fault. Therefore, future decisions regarding *McIntyre*'s effects will presumably involve similarly precise analyses of Tennessee's statutory and common law to ensure that the defendant's interest is not construed too broadly.[166]

---

160. *See id.* at 220–23.

161. *See id.* at 228.

162. *See id.* at 225–26.

163. *Id.* at 220 (citing Brown v. Wal-Mart Disc. Cities, 12 S.W.3d 785, 787 (Tenn. 2000)).

164. *See* Dumas v. State *ex rel.* Dep't of Culture, Recreation & Tourism, 828 So.2d 530, 538 (La. 2002).

165. *See id.*; *see also* Carroll v. Whitney, 29 S.W.3d 14, 19 (Tenn. 2000) (prohibiting the application of joint and several liability despite the fact that the plaintiff's recovery could be diminished or prohibited by the attribution of fault to immune parties).

166. *See, e.g.*, *Banks*, 301 S.W.3d at 221–23; Ali v. Fisher, 145 S.W.3d 557, 561–65 (Tenn. 2004); Owens v. Truckstops of Am., 915 S.W.2d 420, 425–32 (Tenn. 1996).

As *Banks* illustrates, such an analysis may involve extricating the concept of fault from related concepts such as causation.[167] Because the traditional original tortfeasor rule imposed joint and several liability, its drafters had no need to differentiate between the concepts therein: an original tortfeasor both caused and was at fault for subsequent medical negligence.[168] To ensure that applying comparative fault did not unnecessarily impinge on the plaintiff's interest in full recovery, the *Banks* court had to dissect the rule and consider its constituent parts.[169] The comparative fault regime's goal of ensuring proportional liability required the court to disapprove of the rule's second principle—imposing liability disproportional to—as contrary to this goal.[170] However, the court upheld rule's first principle—merely establishing causation—because it did not independently contravene the goal of proportional liability: a defendant could be liable for subsequent injuries but would remain responsible for only the damages attributable to its fault.[171]

The court's specific holding regarding the burden of proof indicates that dissecting procedural rules is also necessary.[172] While Tennessee's modified comparative fault rule permits a court to allocate fault to a nonparty, the regime's underlying policy of proportional liability does not become relevant unless the defendant was not entirely at fault.[173] Until then, the policy of ensuring the plaintiff's full recovery guides procedure; thus, interpreting *McIntyre*'s effect as making such recovery more difficult is impermissible.[174] Despite the well-established rule that the party asserting an affirmative defense bears the burden of proving that defense, some argued that, where the defendant has asserted a comparative fault defense against a nonparty, the defendant was not obliged to prove the nonparty's fault if the plaintiff thereafter amended the complaint to sue the nonparty.[175] The court clarified that the burden remains with the defendant who asserted the defense, and therefore the amending plaintiff will receive full compensation without having to prove the subsequent medical provider's negligence.[176] The court of appeals has subsequently affirmed this holding

167. *See Banks*, 301 S.W.3d at 221, 223.

168. *See* Atkinson v. Hemphill, No. 01-A-01-9311-CV00509, 1994 WL 456349, at *2 (Tenn. Ct. App. Aug. 24, 1994).

169. *See Banks*, 301 S.W.3d at 221–23.

170. *Id.* at 223.

171. *See id.*

172. *See id.* at 225–26.

173. *See id.*

174. *See id.* at 220, 225.

175. *See id.* at 224–25; John A. Day, *Hoping Don't Make It So*, TENN. B.J., May 2010, at 35.

176. *Banks*, 301 S.W.3d at 225.

in opinions regarding the allocation of fault between defendants under the original tortfeasor rule[177] and between a plaintiff and defendant.[178]

B. The Owens Rule and Corollary Will Create Fair Outcomes by Ensuring Proportional Liability

The *Banks* holding suggests that, to achieve the fair balance of plaintiffs' and defendants' interests sought under the comparative fault regime, both the *Owens* rule and its corollary should apply absolutely: comparative fault principles should apply—and thus joint and several liability should be prohibited—where multiple tortfeasors' separate, independent negligent acts combine to cause a single, indivisible injury.[179] However, the *Banks* court's reasoning created unnecessary ambiguity as to the scope and underlying principles of the *Owens* rule and corollary.[180]

The court analyzed relevant post-*McIntyre* cases in light of *McIntyre*'s declaration that the doctrine of joint and several liability was obsolete, but *McIntyre*'s more enduring legacy is its guidance that closely linking liability to fault promotes fairness considerations.[181] In light of exceptions permitting the doctrine's continued application in certain circumstances, mere statements as to the doctrine's obsolescence defy meaningful application.[182] However, these exceptions do evidence the enduring goal of linking liability to fault: each represents circumstances in which the defendant might rightfully be deemed liable for another's conduct, and because liability in these circumstances is not based on the defendant's own negligence, allocation of fault would be inescapably speculative.[183] The facts in *Banks* did not constitute such a circumstance, and thus the court's prohibition of joint and several liability therein did link liability to fault. Nevertheless, given the court's evident goal of providing a broad framework for applying comparative fault principles in future cases, this aim would have been better served by a discussion of post-*McIntyre* precedent in light of linking liability to fault, the common thread among both the cases supporting joint and several liability's obsolescence and the exceptions thereto.[184]

---

177. *See* Stanfield v. Neblett, No. W2009-01891-COA-R3-CV, 2010 WL 2875206, at *2 (Tenn. Ct. App. July 23, 2010) (citing *Banks*, 301 S.W.3d at 224).

178. *See* Wilburn v. City of Memphis, No. W2009-00923-COA-R3-CV, 2010 WL 1407233, at *6 (Tenn. Ct. App. Apr. 9, 2010).

179. *See Banks*, 301 S.W.3d at 219–20, 224, 226, 228 (reiterating the *Owens* rule and the *Owens* corollary).

180. *See id.* at 219.

181. *See id.* at 218 (quoting McIntyre v. Balentine, 833 S.W.2d 52, 58 (Tenn. 1992)).

182. *See id.*

183. *See id.* at 218–20.

184. *See id.*

In addition, the court's failure to discuss the rationale behind these exceptions left unclear the scope of its holding adopting the *Owens* corollary.[185] By merely stating that joint and several liability no longer applies in cases involving separate, independent acts (and listing the exceptions which did not involve such acts), the court's holding forces a question: What constitutes "separate" and "independent" acts?[186] The court in *Ali v. Fisher* suggested that the test is whether a defendant's liability requires a showing of different facts.[187] If this test defines the scope of the *Owens* corollary, the *Banks* court's holding is correct: while Dr. Boyce's liability would depend on evidence of his malpractice, the Elks Lodge's liability for the same injury would turn on evidence of its liability for the original injury. Nevertheless, some reference to this test would have benefited courts subsequently considering the *Banks* court's holding.

Regardless of these ambiguities regarding the rationale and scope of the *Owens* corollary, the *Banks* court's reasoning suggests that it intended to establish the *Owens* corollary as an absolute limitation on the "certain limited circumstances" in which it would permit the continued application of joint and several liability.[188] The court declared that the policy goal in cases under the comparative fault regime has been to strike a fair balance between plaintiffs' and defendants' interests, a goal that is accomplished by linking liability to fault.[189] By stipulating that "the doctrine of joint and several liability no longer applies to circumstances in which separate, independent negligent acts of more than one tortfeasor combine to cause a single, indivisible injury,"[190] the *Owens* corollary necessarily links liability to fault because, in such circumstances, each defendant is only responsible

185. *See id.* at 228.

186. *Id.* While the court also did not discuss the second aspect of the *Owens* corollary, the "single, indivisible injury" provision, this provision does not limit the corollary's scope but merely clarifies that the corollary applies where joint and several liability would have previously applied. *Id.* This provision is, however, a limitation on the scope of the principal *Owens* rule requiring the allocation of fault between plaintiffs and defendants. *See* Mercer v. Vanderbilt Univ., Inc., 134 S.W.3d 121, 128 (Tenn. 2004). However, the *Mercer* decision indicates that the court may not distinguish between indivisible and separate injuries unless such a distinction is obvious. *See id.*

187. Ali v. Fisher, 145 S.W.3d 557, 563 (Tenn. 2004). While not stated expressly, the *Ali* holding and the exceptions permitting joint and several liability cited in *Banks* suggest that the rule's converse is also valid: joint and several liability applies where acts are not separate and independent. *See Banks*, 301 S.W.3d at 219–20; *Ali*, 145 S.W.3d at 563.

188. *See Banks*, 301 S.W.3d at 218 (quoting McIntyre v. Balentine, 833 S.W.2d 52, 60 (Tenn. 1992)).

189. *Id.* at 220 (citing Biscan v. Brown, 160 S.W.3d 462, 474 (Tenn. 2005)); *Ali*, 145 S.W.3d at 563–64; Carroll v. Whitney, 29 S.W.3d 14, 16–17 (Tenn. 2000)).

190. *See id.* at 223 (citing Sherer v. Linginfelter, 29 S.W.3d 451, 455 (Tenn. 2000)); Samuelson v. McMurtry, 962 S.W.2d 473, 476 (Tenn. 1998); Owens v. Truckstops of Am., 915 S.W.2d 420, 430 (Tenn. 1996)).

for its own conduct.[191] By thus ensuring proportional liability, the *Owens* corollary will likewise create the fair balancing of interests sought by the court.[192] Therefore, as long as subsequent courts follow the *Banks* court's example by applying the *Owens* corollary only to the extent necessary to establish proportional liability, this corollary should not be subject to future exceptions based on policy considerations.[193]

This conclusion, drawn from the *Banks* holding, is supported by the fact that no exceptions have been made to the *Owens* corollary.[194] While the *Banks* court's response to the plaintiff's public policy arguments did not directly state that the *Owens* corollary ensured fair outcomes, this response did imply that, as long as the corollary is applied properly—that is, only to the extent necessary to further the goal of proportional liability—it will yield results that are neither burdensome nor costly to plaintiffs.[195]

Despite the court's evident intent to establish the *Owens* corollary as encompassing relevant fairness considerations, the court's failure to reconcile the principal *Owens* rule with the *Mercer* holding's apparent exception thereto might undermine this goal.[196] Granted, the court's holding reiterated only the *Owens* corollary prohibiting joint and several liability, distinct from the principal *Owens* rule requiring allocating fault between plaintiffs and defendants.[197] However, the *Mercer* holding does call into question the reasoning behind the *Banks* court's suggestion that applying the *Owens* corollary will necessarily lead to the proper balance of plaintiffs' and defendants' interests.[198]

While the *Mercer* decision did not expressly create an exception to the *Owens* rule, the *Mercer* court did imply that the rule's function, linking liability with fault, would not necessarily create fair outcomes.[199] Considering circumstances within the scope of the *Owens* rule,[200] the

---

191. *See* Resolution Trust Corp. v. Block, 924 S.W.2d 354, 356 (Tenn. 1996) (holding that when a tort results from multiple individuals' concerted actions, any one of these individuals can be held jointly and severally liable for the other individuals' misconduct).

192. *See Banks*, 301 S.W.3d at 220 (discussing the *Owens* rule as a "core principle" of the comparative fault regime).

193. *See id.* While the circumstances in *Banks* did not provide the court with an opportunity to likewise apply the principal *Owens* rule, this rule will also presumably not be subject to exceptions because it similarly ensures that liability is linked to fault, specifically applying where a plaintiff is comparatively at fault.

194. *See supra* note 75 and accompanying text (listing the exceptions to the doctrine of joint and several liability's general obsolescence, none of which involves separate, independent acts).

195. *See Banks*, 301 S.W.3d at 225–26.

196. *See id.* at 220 (discussing the *Owens* rule as a "core principle" of the comparative fault regime).

197. *See id.* at 228.

198. *See id.* at 220.

199. *See* Mercer v. Vanderbilt Univ., Inc., 134 S.W.3d 121, 129–30 (Tenn. 2004).

200. The *Mercer* court did not dispute that the plaintiff's acts, in negligently causing his

Mercer court held that comparative fault principles should not apply therein based on the policy decision that, where a plaintiff's negligence merely provides the occasion for a subsequent healthcare provider's negligence, the plaintiff should not be denied recovery because he was more than fifty percent at fault for the entire injury.[201] Thus, the *Mercer* court implicitly presented its decision as an exception to the *Owens* rule that, "where the separate, independent negligent acts of more than one tortfeasor combine to cause a single, indivisible injury, each tortfeasor will be liable only for that proportion of the damages attributed to its fault."[202]

While the *Banks* court discussed the *Mercer* holding in the context of clarifying that it was not contrary to the original tortfeasor rule,[203] the court did not address whether *Mercer* represented an exception to the *Owens* rule, cited in the *Banks* opinion as a "core principle" of the comparative fault regime.[204] Such an exception, which prohibits allocating fault to plaintiffs, would not directly affect the *Owens* corollary, since the corollary applies to the allocation of fault among defendants, but it would call into question the rationale behind the *Banks* court's apparent intent to establish the corollary as absolute.[205] As the *Banks* opinion implies, the court adopted a corollary of the *Owens* rule because the *Owens* rule links liability to fault and thereby ensures the court's policy goal of striking a fair balance between plaintiffs' and defendants' interests.[206] However, the *Mercer* holding suggests that, in certain circumstances, striking a fair balance between plaintiffs and defendants' interests may require *not* linking liability to fault and that, in those circumstances, the *Owens* rule will not ensure a fair balancing of interests.[207] In light of this reasoning, both the *Owens* rule and the *Banks* holding adopting the *Owens* corollary will remain subject to exceptions based on fairness considerations where the court reasons that linking liability to fault does not sufficiently address such considerations. Therefore, the *Banks* court may have intended that the *Owens* corollary be absolute, but subsequent courts—like the *Atkinson* court—may cite *Mercer* as permitting exceptions thereto.[208]

---

accident, and the defendant-physician's acts, in negligently rendering treatment thereafter, were separate and independent acts. *See id.* The court held that the distinction between indivisible and separate injuries defied meaningful application therein, but in doing so, failed to distinguish the facts therein from those in *Gray* to which the *Owens* rule applied. *See id.* at 128.

201. *See Banks*, 301 S.W.3d at 226 (discussing *Mercer*, 134 S.W.3d at 129–30).

202. Owens v. Truckstops of Am., 915 S.W.2d 420, 430 (Tenn. 1996).

203. *See Banks*, 301 S.W.3d at 226–27.

204. *Id.* at 220.

205. *See id.*

206. *See id.*

207. *See* Mercer v. Vanderbilt Univ., Inc., 134 S.W.3d 121, 129 (Tenn. 2004).

208. *See* Atkinson v. Hemphill, No. 01-A-01-9311-CV00509, 1994 WL 456349, at *2–3 (Tenn. Ct. App. Aug. 24, 1994).

While the *Banks* court did not address this apparent contradiction between the *Mercer* holding and the *Owens* rule, it might have reconciled the two in light of the original tortfeasor rule. Although the original tortfeasor rule did not apply to the circumstances in *Mercer*, considering the *Mercer* holding in the rule's terms is helpful. Had the plaintiff in *Mercer* been considered the original tortfeasor, the rule would deem his negligence the proximate cause of the damage flowing from his physician's malpractice.[209] Therefore, the court could have apportioned fault for the malpractice to him.[210] By holding otherwise, the *Mercer* rule could be considered the inverse of the original tortfeasor rule, establishing by law that the plaintiff does *not* proximately cause the subsequent medical negligence. This view preserves the *Owens* rule without exception by clarifying that the plaintiff's separate, independent act did not legally *cause* the subsequent injury. Therefore, it would logically follow that, because the plaintiff was not at fault for the subsequent injury, the *Mercer* holding did not deviate from the comparative fault regime's goal of linking liability to fault.

Had *Mercer* adopted this reasoning, the court would thus have been able to avoid conflict with the *Owens* rule. But even if *Mercer* represents an exception to the *Owens* rule, the *Banks* opinion suggests that this holding's scope will be limited to circumstances in which a plaintiff would otherwise forfeit recovery entirely.[211] Therefore, the scope of the *Mercer* holding will likely be limited to such circumstances, and the court will likely reject any argument that *Mercer* stands for the proposition that linking liability to fault will not necessarily strike the proper balance between plaintiffs and defendants' interests. Thus, because the *Mercer* holding restricts merely the allocation of fault to plaintiffs rather than among defendants and because this holding's scope will likely be limited to the facts therein, the *Mercer* holding is unlikely to undermine the *Banks* court's apparent intent to establish the *Owens* corollary as absolute.[212]

---

209. *See* Transports, Inc. v. Perry, 414 S.W.2d 1, 4–5 (Tenn. 1967).

210. *See Banks*, 301 S.W.3d at 228.

211. *See id.* at 226.

212. This conclusion would be far less clear absent the supreme court's numerous reaffirmations that the goal of Tennessee's comparative fault regime is linking liability to fault. *See, e.g.*, Biscan v. Brown, 160 S.W.3d 462, 474 (Tenn. 2005); Ali v. Fisher, 145 S.W.3d 557, 563–64 (Tenn. 2004); Carroll v. Whitney, 29 S.W.3d 14, 21 (Tenn. 2000). Where a comparative fault regime's goal is more limited, courts therein are more likely to adopt exceptions to comparative fault based on policy considerations. *See, e.g.*, Edwards v. Sisler, 691 N.E.2d 1252 (Ind. Ct. App. 1998). Considering circumstances similar to those in *Banks* in a comparative fault system intended merely to modify the "harsh" comparative negligence rule rather than fully link liability to fault, the *Edwards* court retained joint and several liability under the original tortfeasor rule based on, among other factors, the policy consideration that holding otherwise "would interject a stranger into the doctor-patient relationship with an incentive to place the physician and patient at odds" and a motivation to "scour[] an injured party's medical records searching for some act of malpractice . . . to

C. Application of the Original Tortfeasor Rule's First Principle Will Be Unaffected

While the *Banks* holding "did not alter Tennessee's common law rules with regard to liability of tortfeasors for injuries caused by subsequent medical treatment for the injuries they cause [the first principle of the original tortfeasor rule]," the *Banks* court's reasoning in this regard left significant ambiguity as to the first principle's precise basis and scope.[213] The *Banks* holding adopted verbatim the principle as articulated in the Restatement (Third) of Torts: Liability for Physical and Emotional Harm:

> An actor whose tortious conduct is a factual cause of harm to another is subject to liability for any enhanced harm the other suffers due to the efforts of third persons to render aid reasonably required by the other's injury, so long as the enhanced harm arises from a risk that inheres in the effort to render aid.[214]

This version's drafters noted that liability thereunder was a purely legal creation not based on actual foreseeability.[215] While the inherent risks of properly provided treatment are foreseeable, a rule implying that a reasonable person in the original tortfeasor's position would foresee subsequent medical negligence as a natural and probable consequence of its conduct "stretche[d] the idea of probability too far."[216] Nevertheless, the drafters deemed this rule necessary to eliminate the incentive for defendants in enhanced-injury cases to bring malpractice claims to avoid liability.[217]

By contrast, the *Banks* court reasoned that this principle was based on the rationale that liability for successive injuries arises when "the subsequent negligent conduct is a foreseeable or natural consequence of the original tortfeasor's negligence."[218] Thus, the *Banks* court considered the first principle as merely clarifying a negligence claim's proximate cause

---

extricate himself from all or some portion of the damages to the injured party." *Id.* at 1255 (citing Holden v. Balko, 949 F. Supp. 704, 711–12 (S.D. Ind. 1996)).

213. *Banks*, 301 S.W.3d at 223.

214. RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL & EMOTIONAL HARM § 35 (2010); *see Banks*, 301 S.W.3d at 222.

215. *See* RESTATEMENT (SECOND) OF TORTS § 457 cmt. b (1965), *reiterated in* RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL & EMOTIONAL HARM § 35 cmt. b (2010).

216. RESTATEMENT (SECOND) OF TORTS § 457 cmt. b (1965), *reiterated in* RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL & EMOTIONAL HARM § 35 cmt. b (2010). While the latter does not include this same language, it implies that subsequent medical negligence is not "plainly within the scope of the actor's liability." *Id.*

217. RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL & EMOTIONAL HARM § 35 cmt. b (2010).

218. *Banks*, 301 S.W.3d at 222.

element rather than implying liability by law.[219] Given the explicit wording of the version the *Banks* court adopted, these differences in rationale should not bear on the first principle's future functionality, but they do clarify that, despite apparent differences between this version and the prior common law version, the *Banks* holding did not change the first principle's scope.[220]

While the common law version did not contain the express limitations that subsequent injuries must have occurred during treatment that the original injury "reasonably required" and must have arisen from "a risk that inheres in the effort to render aid," this version's basis in foreseeability suggests that these considerations were nevertheless relevant thereunder.[221] While the court deemed subsequent medical negligence foreseeable based on the "predictable incidence of medical malpractice," an original tortfeasor has no reason to foresee unnecessary treatments or harms that the "normal efforts to render aid" did not create.[222] Conversely, the version that the *Banks* court adopted included these express provisions based on the rationale that, because the original tortfeasor's liability for subsequent medical negligence is an artificial construct, foreseeability limitations are irrelevant unless expressly stated therein. Therefore, the semantic differences between the *Banks* court's version of the rule and the prior common law version do not indicate that the *Banks* holding changed the scope of the original tortfeasor rule's first principle.

VI. CONCLUSION

The Tennessee Supreme Court's holding in *Banks* reflects the court's ongoing reassessment of statutory and common law rules in light of the comparative fault principles adopted eighteen years earlier in *McIntyre*.[223] Although the court's post-*McIntyre* opinions have evidenced some dispute over whether the joint and several liability doctrine is "obsolete," this conflict is merely semantic: these holdings illustrate the common goal of linking liability to fault.[224] The *Banks* holding is particularly instructive regarding the manner in which a common law rule should be dissected and analyzed to ensure that the plaintiff's interest in full recovery is abridged only where the provisions pursuant to this interest are irreconcilable with the goal of proportional liability.[225] While the court's procedural holding

219. *See* Naifeh v. Valley Forge Life Ins. Co., 204 S.W.3d 758, 771 (Tenn. 2006) (outlining the elements of negligence).

220. *See* Transports, Inc. v. Perry, 414 S.W.2d 1, 4–5 (Tenn. 1967).

221. *See id.* at 4–6 (considering whether the injury that the original tortfeasor caused actually necessitated the subsequent medical treatment).

222. RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL & EMOTIONAL HARM § 35 cmt. b & c (2010).

223. *See Banks*, 301 S.W.3d at 218–20.

224. *See supra* notes 181–184 and accompanying text.

225. *See supra* notes 163–171 and accompanying text.

has had an immediate impact,[226] the long-term ramifications of this decision will likely be much broader. Both the opinion's thorough review of post-*McIntyre* precedent and its coalescence of these holdings into broader principles indicate that the *Banks* court intended to create a clear framework for applying comparative fault in a wide range of future cases.[227] Despite a rationale that is at times ambiguous, the court's holding should nevertheless establish more finite parameters as to the tenets and goals of the comparative fault regime.

DAN CALVERT

226. *See supra* notes 172–178 and accompanying text.
227. *See supra* notes 188–95 and accompanying text.

BANKRUPTCY LAW—DISCHARGEABILITY OF STUDENT LOAN DEBT—APPLICATION OF FEDERAL RULE OF CIVIL PROCEDURE 60(B)(4) TO DISCHARGE-BY-DECLARATION PROVISIONS

United Student Aid Funds, Inc. v. Espinosa, 130 S. Ct. 1367 (2010).

I. INTRODUCTION

To finance his education, respondent Francisco Espinosa ("Espinosa") executed four student loan promissory notes between 1988 and 1989 for a total original principal amount of $13,250.[1] The Federal Family Education Loan Program administered and federally guaranteed the educational loans.[2] In 1992, Espinosa filed a petition and plan for relief under Chapter 13[3] of the United States Bankruptcy Code.[4] Under Chapter 13, "[a] proposed bankruptcy plan becomes effective upon confirmation . . . and will result in a discharge of the debts listed in the plan if the debtor completes the payments the plan requires."[5]

Espinosa's bankruptcy plan proposed the repayment of only the principal portion of his student loan obligation to petitioner United Student Aid Funds, Inc. ("United"), the guarantor and holder of the notes.[6] The plan also provided that any "accrued capitalized interest, penalties[,] and fees" would be discharged once Espinosa had paid in full the amount of $13,250

1. United Student Aid Funds, Inc. v. Espinosa, 130 S. Ct. 1367, 1373 (2010).

2. Brief for Petitioner at 2, United Student Aid Funds, Inc. v. Espinosa, 130 S. Ct. 1367 (2010) (No. 08-1134). The Federal Family Education Loan Program ("FFELP"), managed by the U.S. Department of Education, was established by Congress as part of the Higher Education Act of 1965, now codified at 20 U.S.C. §§ 1001–1155 (2006). *See* Chae v. SLM Corp., 593 F.3d 936, 938 (9th Cir. 2010) (describing the FFELP as "a system of loan guarantees meant to encourage [private] lenders to loan money to students and their parents on favorable terms").

3. "Under Chapter 13 of the Bankruptcy Code, individual debtors may obtain adjustment of their indebtedness through a flexible repayment plan approved by a bankruptcy court." Nobelman v. Am. Sav. Bank, 508 U.S. 324, 327 (1993). "[T]he elements of a confirmable Chapter 13 plan" are provided under the provisions of 11 U.S.C. § 1322. *Id.*

4. *Espinosa*, 130 S. Ct. at 1373.

5. *Id.* (citing 11 U.S.C. §§ 1324, 1325, 1328(a)) (internal citations omitted).

6. *Id.* at 1374. "United is a guaranty agency that administers the collection of federally guaranteed student loans in accordance with regulations promulgated by the United States Department of Education." *Id.* at 1374 n.1. One of the transactional functions of the FFELP involves regulating guaranty agencies. *See Chae*, 593 F.3d at 939 ("A guaranty agency is a 'State or private nonprofit organization that has an agreement with the Secretary [of the Department of Education] under which it will administer a loan guarantee program under the [Higher Education] Act.'") (quoting 34 C.F.R. § 682.200(b)(2)(vi))).

as unsecured claims.[7] After receiving a notice of commencement and a copy of Espinosa's proposed plan from the Bankruptcy Court, United responded by filing a "proof of claim" in this case for $17,832.15—an amount constituting both the principal and pre-petition interest on the respondent's student loans.[8] Interestingly, United, as the only creditor specified, did not file any objections to Espinosa's proposed Chapter 13 plan.[9] In 1993, the Bankruptcy Court issued a confirmation order, approving Espinosa's Chapter 13 protection plan in the absence of a judicial declaration of "undue hardship" in an adversary proceeding.[10] United did not appeal the confirmation order, nor did it respond one month

---

7. Espinosa v. United Student Aid Funds, Inc., 530 F.3d 895, 896 (9th Cir. 2008) (per curiam) (holding that the bankruptcy court properly ordered the creditor to discontinue all collection actions against the debtor until the court could determine on remand whether a clerical error had been introduced into the discharge order and should be corrected). In boldface type on the first page, the debtor's Chapter 13 Plan stated this admonition: "WARNING IF YOU ARE A CREDITOR YOUR RIGHTS MAY BE IMPAIRED BY THIS PLAN." Brief of Appellant at 8–9, Espinosa v. United Student Aid Funds, Inc., 530 F.3d 895 (9th Cir. 2008) (No. 06-16421). Espinosa's plan also expressly declared that "[a]ny amounts or claims for student loans unpaid by this Plan shall be discharged." *Id.* at 10.

8. *Espinosa*, 130 S. Ct. at 1374; *Espinosa*, 530 F.3d at 896.

9. *Espinosa*, 130 S. Ct. at 1374.

10. *Id.* Under the Bankruptcy Code, educational loan debts are only dischargeable upon a judicial finding that failure to discharge such debt would inflict "undue hardship" on the debtor and his or her dependents. 11 U.S.C. §§ 523(a)(8), 1328 (2006). Section 523(a), in pertinent part, states:

> A discharge under section 727, 1141, 1228(a), 1228(b), or 1328(b) of this title does not discharge an individual debtor from any debt–
>
> . . . .
>
> (8) unless excepting such debt from discharge under this paragraph would impose an undue hardship on the debtor and the debtor's dependents, for–
>
> (A)(i) an educational benefit overpayment or loan made, insured, or guaranteed by a governmental unit, or made under any program funded in whole or in part by a governmental unit or nonprofit institution; or (ii) an obligation to repay funds received as an educational benefit, scholarship, or stipend; or
>
> (B) any other educational loan that is a qualified education loan, as defined in section 221(d)(1) of the Internal Revenue Code of 1986, incurred by a debtor who is an individual . . .

11 U.S.C. § 523(a)(8) (2006). Further circumscribing the dischargeability of student loan debt, the Federal Rules of Bankruptcy Procedure require the debtor to establish "undue hardship" in an adversary proceeding. FED. R. BANKR. P. 7001(6). To initiate an adversary proceeding, the debtor must serve a complaint and summons on the creditor-defendant. FED. R. BANKR. P. 7003, 7004.

later after being served notice of the Chapter 13 trustee's objection to United's proof of claim.[11]

In 1997, Espinosa completed payment of his student loan obligations to United as stipulated in his Chapter 13 plan, and the Bankruptcy Court subsequently discharged the outstanding interest remaining on Espinosa's student loan debt.[12] Three years after the bankruptcy court issued the discharge order, the U.S. Department of Education[13] began seizing Espinosa's federal income tax refunds in an attempt to recoup the unpaid portion of the respondent's student loan debt.[14] Consequently, in 2003, Espinosa petitioned the Bankruptcy Court for an order "enforc[ing] its 1997 discharge order" by enjoining both the Department and United from further collection activities.[15] In response, United filed a cross-motion for relief on the basis that the Bankruptcy Court's 1993 confirmation order was void under Federal Rule of Civil Procedure 60(b)(4).[16] United's Rule 60(b)(4)

---

11. Brief for Respondent at 4–5, United Student Aid Funds, Inc. v. Espinosa, 130 S. Ct. 1367 (2010) (No. 08-1134). "The [Chapter 13 Trustee's] objection notified United that its claim of $17,832.15 would only be paid in the amount of $13,250.00. The Trustee's notice provided thirty days to object to this treatment of United's claim. United did not object or appeal." *Id.*

12. *Espinosa*, 130 S. Ct. at 1374. The Supreme Court noted that "[t]he one-page discharge order contained a paragraph that purported to *exclude 'any debt . . . for a student loan'* from the discharge." *Id.* at 1375 n.4 (quoting *Espinosa*, 530 F.3d at 896) (emphasis added). This glaring inconsistency between the language of the discharge order and the terms of Espinosa's Chapter 13 plan was eventually addressed when the United States Court of Appeals for the Ninth Circuit remanded the case to the Bankruptcy Court for this purpose. *Espinosa*, 530 F.3d at 899. Pursuant to FED. R. CIV. P. 60(a), the Bankruptcy Court found the contradictory text in the discharge order to be the result of a "clerical mistake" and expunged the language from the order. *Espinosa*, 130 S. Ct. at 1374 n.4.

13. The Supreme Court clarified: "After Espinosa completed payments under the plan, United assigned Espinosa's loans to the Department [of Education] under a reinsurance agreement. . . . . [In 2004], United requested and received a recall of the loans from the Department." *United*, 130 S. Ct. at 1374 n.3 (citation omitted); *see* Brief for Petitioner at 9, United Student Aid Funds, Inc. v. Espinosa, 130 S. Ct. 1367 (2010) (No. 08-1134).

14. *Espinosa*, 130 S. Ct. at 1374.

15. *Id.*

16. *Id.* FED. R. CIV. P. 60(b) provides:

> On motion and just terms, the court may relieve a party or its legal representative from a final judgment, order, or proceeding for the following reasons:
>
> (1) mistake, inadvertence, surprise, or excusable neglect;
> (2) newly discovered evidence that, with reasonable diligence, could not have been discovered in time to move for a new trial under Rule 59(b);
> (3) fraud (whether previously called intrinsic or extrinsic), misrepresentation, or misconduct by an opposing party;
> (4) *the judgment is void*;

motion turned on two arguments: (1) that the order confirming Espinosa's plan violated United's rights under the Bankruptcy Code and Rules by not requiring a judicial finding of undue hardship in an adversary proceeding; and (2) that United's due process rights were violated because of Espinosa's failure to fulfill the specific service of process requirements prior to an adversarial proceeding.[17]

The United States Bankruptcy Court for the District of Arizona found United's arguments unpersuasive and denied its motion asking that the confirmation order be set aside as void.[18] Furthermore, the Bankruptcy Court "granted Espinosa's motion in relevant part" and instructed Espinosa's creditors to desist from all collection efforts against Espinosa.[19] On appeal, the United States District Court for the District of Arizona reversed, holding that the Bankruptcy Court's confirmation order was void because United was denied due process notice.[20] Espinosa appealed the District Court's ruling.[21] In its per curiam opinion, the United States Court of Appeals for the Ninth Circuit initially remanded the case to the Bankruptcy Court for the limited purpose of permitting the lower court to consider correcting a possible oversight or "clerical error" in the discharge order for Espinosa's student loan debt.[22] Following correction of the clerical mistake by the Bankruptcy Court and resubmission of the case, a unanimous Ninth Circuit reversed the District Court's judgment, concluding that Espinosa's failure to comply with procedural prescriptions under the Bankruptcy Code and Rules did not render, *ipso facto*, a court's discharge order void.[23] The Ninth Circuit also found that United had received constitutionally adequate notice under the particular circumstances but simply failed to object to Espinosa's proposed Chapter 13 plan.[24] On certiorari to the United States Supreme Court, *held*, affirmed.[25] A bankruptcy court's confirmation order adopting a debtor's plan for discharging a student loan debt in the absence of an undue hardship determination or adversary proceeding is not a void judgment under Rule

---

> (5) the judgment has been satisfied, released or discharged; it is based on an earlier judgment that has been reversed or vacated; or applying it prospectively is no longer equitable; or
> (6) any other reason that justifies relief.

FED. R. CIV. P. 60(b) (emphasis added).

17. *Espinosa*, 130 S. Ct. at 1374–75.

18. *Id.* at 1375.

19. *Id.*

20. Espinosa v. United Student Aid Funds, Inc., No. CV-04-447-TUC-RCC, 2006 WL 6296057, at *4 (D. Ariz. July 18, 2006).

21. *Espinosa*, 130 S. Ct. at 1375.

22. Espinosa v. United Student Aid Funds, Inc., 530 F.3d 895, 899 (9th Cir. 2008).

23. Espinosa v. United Student Aid Funds, Inc., 553 F.3d 1193, 1198 (9th Cir. 2008).

24. *Id.* at 1203.

25. *Espinosa*, 130 S. Ct. at 1382.

60(b)(4), because (1) deprivation of "a right granted by a procedural rule" is not tantamount to a violation of a claimant's "constitutional right to due process," and (2) a bankruptcy court's failure to find undue hardship constituted a "legal error," which does not entitle a claimant to relief under Rule 60(b)(4).[26] *United Student Aid Funds, Inc. v. Espinosa*, 130 S. Ct. 1367 (2010).

II. JUDICIAL CONFIRMATION OF STUDENT LOAN DISCHARGE PLANS: STATUTORY AND CONSTITUTIONAL CHALLENGES

With the passage of the Higher Education Act of 1965,[27] the availability of government-guaranteed student loans enabled many individuals, who otherwise lacked the financial resources or credit to qualify for private loans, to pursue postsecondary education.[28] They were considered "enabling loans" because borrowers used the money "to improve their own human capital" and thus increase their potential income.[29] Congress surmised that student borrowers would use their increased future earnings to repay their loan obligations.[30] Based on this congressional view, student loan debt is "presumptively nondischargeable in bankruptcy proceedings"[31] unless the debtor can affirmatively show "undue hardship" in an adversary proceeding.[32]

In the past decade, a clear schism has emerged among the federal appeals courts concerning the enforceability of confirmed Chapter 13 plans for discharging student loans. On the one hand, the Tenth Circuit held that the res judicata effect of confirmed Chapter 13 plans and the "strong policy of finality"[33] supersede any inconsistencies between the court's judgment and the statutory provisions of the Bankruptcy Code.[34] On the other hand, the Fourth, Sixth, and Seventh Circuits have held that confirmation or discharge orders contrary to statutory provisions are void or not entitled to preclusive effect based on a violation of due process that deprives a creditor

---

26. *Id.* at 1378, 1380.

27. *See* Seth J. Gerson, *Separate Classification of Student Loans in Chapter* 13, 73 WASH. U. L.Q. 269, 279–80 (1995).

28. *Id.* at 280.

29. Kevin C. Driscoll Jr., *Eradicating the "Discharge By Declaration" for Student Loan Debt in Chapter 13*, 2000 U. ILL. L. REV. 1311, 1315 (2000).

30. Gerson, *supra* note 27, at 280.

31. *In re* Hanson, 397 F.3d 482, 484 (7th Cir. 2005).

32. 11 U.S.C. § 523(a)(8) (2006).

33. Andersen v. UNIPAC-NEBHELP (*In re* Andersen), 179 F.3d 1253, 1259 (10th Cir. 1999) (affirming Bankruptcy Appellate Panel's prodebtor decision based on the "res judicata effect" of the confirmed Chapter 13 plan and "strong policy favoring finality," despite the fact that the debtor in this case had failed to prove undue hardship).

34. *See id.* at 1258; Great Lakes Higher Educ. Corp. v. Pardee (*In re* Pardee), 193 F.3d 1083, 1087 (9th Cir. 1999) (finding "no reason to depart from the well-settled policy that confirmation orders are final orders that are given preclusive effect").

of notice.[35] Given this manifest inconsistency in bankruptcy adjudications regarding student loan debt, the Supreme Court granted certiorari to resolve the question of "whether an order that confirms the discharge of a student loan debt in the absence of an undue hardship finding or an adversary proceeding, or both, is a void judgment for Rule 60(b)(4) purposes."[36] In *United Student Aid Funds, Inc. v. Espinosa*,[37] the U.S. Supreme Court held that judicial confirmation of a debtor's educational loan adjustment plan without the prerequisite finding of undue hardship in an adversary proceeding is not rendered void under Federal Rule of Civil Procedure 60(b)(4).[38]

A. Statutory Framework for Discharging Student Loan Debt

Over the past thirty-two years, beginning with the enactment of the Bankruptcy Reform Act of 1978,[39] Congress has revised bankruptcy law as it pertains to the discharge of educational loan debts on a number of occasions, gradually mandating such debt from being "readily dischargeable to presumptively nondischargeable."[40] This progressive

---

35. *See In re Hanson*, 397 F.3d at 487 ("Due to the lack of compliance with the Bankruptcy Code and Rules, the bankruptcy discharge order was void and [the student loan creditor] was properly granted relief pursuant to Rule 60(b)(4)."); Ruehle v. Educ. Credit Mgmt. Corp. (*In re* Ruehle), 412 F.3d 679, 684 (6th Cir. 2005) ("The [finality analysis in *Andersen* and *Pardee*] embodies many of the dangers inherent in winking at due process, which is the cornerstone of justice.") (alteration in original); Banks v. Sallie Mae Servicing Corp. (*In re* Banks), 299 F.3d 296, 303 (4th Cir. 2002) ("For lack of adequate notice, the confirmation and discharge orders discharging the interest are not entitled to preclusive effect.").

36. *Espinosa*, 130 S. Ct. at 1373.

37. 130 S. Ct. 1367 (2010).

38. *Id.* at 1380.

39. Gerson, *supra* note 27, at 282 (citing Bankruptcy Reform Act of 1978, Pub. L. No. 95-598, 92 Stat. 2549). The Act, as provided in a modified version of § 523(a)(8), mandated a waiting period of five years during which a student loan could not be discharged in a Chapter 7 bankruptcy action unless the restriction created an undue hardship on the debtor. *Id.* at 282 n.101. "[T]his provision was not originally applicable to Chapter 13 cases." *Id.* at 282. In 1990, Congress extended the statutory waiting period to seven years, making it increasingly more difficult for student borrowers to evade their loan obligations. *Id.* at 282 n.101 (citing Pub. L. No. 101-647, 104 Stat. 4789, 4964–65 (1990)). The "loophole" available through the "more liberal discharge provisions of Chapter 13" was closed in 1990 when Congress amended § 1328(a)(2), effectively codifying educational loans as nondischargeable debts. *Id.* at 282–83. Finally, in 1998, Congress again revised section 523 of the Bankruptcy Code, thereby extinguishing a debtor's option to discharge a student loan debt even after the seven-year waiting period. Mersmann v. United Student Aid Funds, Inc. (*In re* Mersmann), 505 F.3d 1033, 1042 (10th Cir. 2007) (citing Higher Education Amendments of 1998, Pub. L. No. 105-244, Title IX, § 971(a), 112 Stat. 1581, 1837 (1998)).

40. Driscoll, *supra* note 29, at 1316. *See also* Gerson, *supra* note 27, at 279–83 (providing an excellent comprehensive treatment of the legislative history of student loan

foreclosure on the dischargeability of student loan debt underscores Congress's intention that such loans should be fully repaid.[41] As one legal commentator noted, "Congress was troubled by the thought of a college student who incurs government-backed loans and wantonly discharges the debts moments after graduation, brazenly embarking on a lifetime of enhanced earnings unfettered by loan payments."[42]

As set forth in 11 U.S.C. § 523(a)(8), which "establishes the statutory presumption against the discharge of student loans,"[43] a debtor may be relieved of the burden of repaying an educational loan debt only under one narrow condition: when "excepting such debt from discharge . . . would impose an *undue hardship* on the debtor and the debtor's dependents . . . ."[44] The Bankruptcy Code does not specifically define "undue hardship"[45] but leaves an interpretation of this evidentiary element to the discretion of the courts. Accordingly, the Second Circuit in *In re Brunner* promulgated a three-pronged test, proposed by the district court, for establishing undue hardship.[46] Under the *Brunner* test, the debtor must affirmatively show:

> (1) that the debtor cannot maintain, based on current income and expenses, a "minimal" standard of living for [himself or] herself and [his or] her dependents if forced to repay the loans; (2) that additional circumstances exist indicating that this state of affairs is likely to persist for a significant portion of the repayment period of the student loans; and (3) that the debtor has made good faith efforts to repay the loans.[47]

However, the Eighth Circuit rejected the *Brunner* test, adopting instead a less restrictive "totality-of-the-circumstances approach to the 'undue hardship' inquiry."[48] The court in *In re Long* cited policy arguments of

---

debts in bankruptcy).

41. Gerson, *supra* note 27, at 283.

42. Driscoll, *supra* note 29, at 1316 (citing H.R. Doc. No. 93-137, pts. I & II (1973)).

43. Whelton v. Educ. Credit Mgmt. Corp. (*In re* Whelton), 432 F.3d 150, 153 (2d Cir. 2005).

44. 11 U.S.C. § 523(a)(8) (2006) (emphasis added). *See supra* note 10 for the complete, relevant portion of the statute.

45. Andrews v. S.D. Student Loan Assistance Corp. (*In re* Andrews), 661 F.2d 702, 704 (8th Cir. 1981) (quoting Wegfehrt v. Ohio Student Loan Comm'n (*In re* Wegfehrt), 10 B.R. 826, 830 (Bankr. N.D. Ohio 1981)).

46. Brunner v. N.Y. State Higher Educ. Servs. Corp. (*In re* Brunner), 831 F.2d 395, 396 (2d Cir. 1987).

47. *Id.* The debtor must prove all three elements of the *Brunner* test by a preponderance of the evidence in order for the Bankruptcy Court to grant a discharge of a student loan debt. *See* 8B C.J.S. *Bankruptcy* § 1078 (2006).

48. Long v. Educ. Credit Mgmt. Corp. (*In re* Long), 322 F.3d 549, 554 (8th Cir. 2003); *see also In re Andrews*, 661 F.2d at 704–05.

"fairness and equity" for its decision "to embrace" the entirety of the facts and circumstances surrounding the particular bankruptcy case at bar.[49]

In *Tennessee Student Assistance Corp. v. Hood*,[50] the Supreme Court declared that the student loan nondischargeability provision, § 523(a)(8), is "self-executing" in the sense that a discharge of a student loan debt will not be granted "[u]nless the debtor affirmatively secures a hardship determination."[51] In accordance with the Federal Rules of Bankruptcy Procedure ("Bankruptcy Rules"), the debtor is required to establish undue hardship in an adversary proceeding.[52] Adversary proceedings that have been termed "subactions"[53] or "minitrial[s],"[54] are initiated as a part of the overall bankruptcy case and possess the accoutrements of the traditional civil litigation process.[55] For example, the debtor initiates an adversary proceeding by first filing a complaint,[56] which is served with an accompanying summons[57] on the creditor. The Bankruptcy Rules require that the creditor-defendant serve a responsive pleading "within 30 days after the issuance of the summons" or within 35 days if the creditor is the United States government.[58]

---

49. *In re Long*, 322 F.3d at 554.

50. Tenn. Student Assistance Corp. v. Hood, 541 U.S. 440, 443 (2004) (holding, in part, that an adversary proceeding initiated by a debtor seeking an undue hardship determination does not constitute "a suit against the State for purposes of the Eleventh Amendment").

51. *Id.* at 450.

52. FED. R. BANKR. P. 7001(6) (defining "a proceeding to determine the dischargeability of a debt" as a type of adversary proceeding).

53. Blevins Elec., Inc. v. First Am. Nat'l Bank (*In re* Blevins Elec., Inc.), 185 B.R. 250, 253 (Bankr. E.D. Tenn. 1995).

54. Driscoll, *supra* note 29, at 1317 (describing the adversary proceeding for an undue hardship determination).

55. *In re* Mersmann, 505 F.3d 1033, 1043 (10th Cir. 2007). "Like a civil trial, adversary proceedings (1) are governed by discovery rules, (2) can be adjudicated by summary judgment, (3) are subject to the award of costs, and (4) are appealable." *Id.* (internal citations omitted).

56. *See* FED. R. BANKR. P. 7003 (applying FED. R. CIV. P. 3 to adversary proceedings in bankruptcy).

57. *See* FED. R. BANKR. P. 7004 (applying the mechanics of service of process as specified in FED. R. CIV. P. 4 to adversary proceedings in bankruptcy).

58. *See* FED. R. BANKR. P. 7012(a). Section 7012(b) states:

> A responsive pleading shall admit or deny an allegation that the proceeding is core or non-core. If the response is that the proceeding is non-core, it shall include a statement that the party does or does not consent to entry of final orders or judgment by the bankruptcy judge. In non-core proceedings final orders and judgments shall not be entered on the bankruptcy judge's order except with the express consent of the parties.

FED. R. BANKR. P. 7012(b).

A necessary step toward the debtor's financial absolution is the filing of a Chapter 13 bankruptcy plan[59]—essentially "a laundry list of whom is to be paid, how much they are to be paid, and over what length of time this payment will occur."[60] The bankruptcy court judge then schedules a hearing to consider confirmation of the debtor's Chapter 13 plan.[61] Generally, notice[62] to the student-loan holder(s) in interest is given by mailing a copy of the debt adjustment plan and notice of the confirmation hearing to the addresses provided in the debtor's list of creditors.[63] At this point, the creditor can object to judicial confirmation of the plan in order to safeguard its rights.[64] The court may approve a Chapter 13 plan provided that the plan complies with the statutory provisions set forth in § 1325 of the Bankruptcy Code and that the holder of any unsecured claim does not object to the confirmation of the plan.[65]

Certain federal appeals courts, however, have affirmed the confirmation of proposed Chapter 13 plans, despite the fact that these plans incorporated provisions explicitly purporting to discharge nondischargeable educational loan debts in the absence of an adversary proceeding.[66] This practice of attempting to circumvent procedural requirements through the inclusion of language that expressly violates the Bankruptcy Code is referred to as "discharge by declaration"[67] and is illustrated in this excerpt from one debtor's Chapter 13 confirmation plan: "Pursuant to 11 U.S.C. § 523(a)(8), excepting . . . education loans from discharge will impose an undue hardship on the debtor and the debtor's dependents. Confirmation of debtor's plan shall constitute a finding to that effect and that said debt is dischargeable."[68] As set forth in § 1327(a), confirmation of a bankruptcy plan is binding on the parties unless appealed or revoked.[69] Not

---

59. 11 U.S.C. § 1321 (2006) ("The debtor shall file a plan.").

60. Driscoll, *supra* note 29, at 1319 (summarizing the essential contents of a Chapter 13 plan as set forth in 11 U.S.C. § 1322(a) (2006)).

61. *See* 11 U.S.C. § 1324(a) (2006).

62. *See* FED. R. BANKR. P. 2002(b) ("[T]he clerk . . . shall give . . . all creditors and indenture trustees not less than 25 days' notice by mail of . . . the time fixed for filing objections and the hearing to consider confirmation of a chapter 9, chapter 11, or chapter 13 plan.").

63. *See id.* ("The plan or a summary of the plan shall be included with each notice of the hearing on confirmation mailed pursuant to Rule 2002.").

64. *See* 11 U.S.C. § 1324(a) (2006).

65. *See id.* § 1325(a)–(b).

66. *See, e.g.*, *In re* Andersen, 179 F.3d 1253, 1260 (10th Cir. 1999); *In re* Pardee, 193 F.3d 1083, 1084 (9th Cir. 1999).

67. *See In re* Mersmann, 505 F.3d 1033, 1038 (10th Cir. 2007); Driscoll, *supra* note 29, at 1320.

68. *In re Andersen*, 179 F.3d at 1254.

69. *See* 11 U.S.C. § 1327(a) (2006) (providing that "[t]he provisions of a confirmed plan bind the debtor and each creditor, whether or not the claim of such creditor is provided for by the plan, and whether or not such creditor has objected to, has accepted, or has

surprisingly, confirmed reorganization plans containing illegal discharge-by-declaration provisions have been challenged collaterally by disgruntled creditors on grounds that such approved provisions are contrary to the Bankruptcy Code[70] and contravene due process rights.[71] The case law indicates a split of circuit court authority as to whether the confirmation of student loan adjustment plans approved in the absence of an adversary proceeding to determine undue hardship is void and therefore unenforceable.[72]

B. Statutory Noncompliance Challenges

In 1999, the Tenth Circuit decided *In re Andersen*, which started a succession of (ultimately) conflicting case law focused on educational loan discharges without proof of hardship.[73] In that case, the Tenth Circuit affirmed the Bankruptcy Appellate Panel's ruling that confirmation of Andersen's Chapter 13 plan "constituted a finding of undue hardship, rendering the student loans dischargeable," despite the fact that Andersen never initiated an adversary proceeding as required by statute.[74] The Chapter 13 plan in question provided that Andersen would pay back ten percent of each student loan claim, at which point an inserted discharge-by-declaration provision released Andersen from the remaining ninety percent upon completion of the plan.[75] Inexplicably, the guarantor of Andersen's

---

rejected the plan."). *See also* Driscoll, *supra* note 29, at 1320.

70. *See In re* Stevens, 236 B.R. 350, 351 (Bankr. E.D. Va. 1999); *In re* Mammel, 221 B.R. 238, 239 (Bankr. N.D. Iowa 1998).

71. *See In re* Hanson, 397 F.3d 482, 484 (7th Cir. 2005); *In re* Ruehle, 412 F.3d 679, 681 (6th Cir. 2005); *In re* Banks, 299 F.3d 296, 299–300 (4th Cir. 2002); Driscoll, *supra* note 29, at 1326.

72. *See In re Mersmann*, 505 F.3d at 1044; Farris E. Ain, *Never Judge a Bankruptcy Plan by its Cover: The Discharge of Student Loans Through Provisions in a Chapter 13 Plan*, 32 SW. U. L. REV. 703, 705 (2003) (discussing the legal arguments and policies proffered in support or rejection of student loan discharge provisions in Chapter 13 plans).

73. *In re Hanson*, 397 F.3d at 485; *accord In re* Andersen, 179 F.3d 1253 (10th Cir. 1999).

74. *In re* Andersen, 179 F.3d at 1255 (citing *In re* Andersen, 215 B.R. 792, 796 (B.A.P. 10th Cir. 1998)).

75. *Id.* at 1254. Seeking an adjustment of her educational loan obligations, the debtor in this case filed a Chapter 13 plan that incorporated statutory language into a discharge-by-declaration provision as follows:

> All timely filed and allowed unsecured claims, including the claims of Higher Education Assistance Foundation and UNIPAC-NEBHELP, which are government guaranteed education loans, shall be paid ten percent (10%) of each claim, and the balance of each claim shall be discharged. Pursuant to 11 U.S.C. § 523(a)(8), excepting the aforementioned education loans from discharge will impose an undue hardship on the debtor and the debtor's

student loans failed to timely object to the proposed plan prior to confirmation or appeal the subsequent confirmation order.[76]

In its analysis, the Tenth Circuit stated that the "real issue" before the court was "whether confirmation of the plan constitute[d] a binding adjudication of hardship"—not whether the debtor had adequately proven the evidentiary element of undue hardship, an issue that, as the Court pointed out, was indisputable, because Andersen clearly had not fulfilled the statutory requirement.[77] On appeal, the creditor in *Andersen* argued that the Bankruptcy Court had exceeded its authority by confirming the debtor's Chapter 13 plan, because a plan containing provisions which are clearly irreconcilable with the Bankruptcy Code is not confirmable.[78] The Tenth Circuit, however, found the creditor's challenge on grounds of statutory noncompliance to be unavailing.[79] As the court emphasized repeatedly throughout its opinion, the student-loan creditor had an affirmative responsibility of protecting its interests by challenging the inconsistent language in the proposed plan and appealing the confirmation order: "A creditor cannot simply sit on its rights and expect that the bankruptcy court or trustee will assume the duty of protecting its interests."[80] From the Tenth Circuit's perspective, the necessity for finality in judicial matters eclipses a court's obligation to correct orders that erroneously confirm plans which are at variance with the Bankruptcy Code.[81] In its holding, therefore, the Tenth Circuit cited the "compelling need for finality" of judgments and the doctrine of res judicata as strong arguments for discharging the outstanding balance on Andersen's student loan debt in accordance with the confirmation and discharge orders.[82]

The Ninth Circuit in *In re Pardee* espoused the line of reasoning articulated by the Tenth Circuit in *Andersen*.[83] In *Pardee*, the confirmed Chapter 13 plan at issue claimed to discharge the debtors' post-petition interest on a student loan debt.[84] Like the claimant in *Andersen*, the creditor-defendant in *Pardee* failed to raise an objection to the proposed plan or to appeal the confirmation order.[85] The creditor argued on appeal to the Bankruptcy Court's discharge order that the post-petition interest on the

---

> dependents. Confirmation of debtor's plan shall constitute a finding to that effect and that said debt is dischargeable.

Id.

76. *Id.*
77. *Id.* at 1256.
78. *Id.* at 1257.
79. *Id.*
80. *Id.*
81. *Id.* at 1258.
82. *Id.* at 1258–59.
83. *In re* Pardee, 193 F.3d 1083, 1086 (9th Cir. 1999).
84. *Id.* at 1084.
85. *Id.*

education loan was nondischargeable pursuant to §§ 523(a)(8) and 1328(a)(2);[86] confirmation of a plan purporting to discharge such interest, therefore, violated statutory requirements.[87] Invoking the rationale of the Tenth Circuit, the Ninth Circuit stated that "[i]f a creditor fails to protect its interests by timely objecting to a plan or appealing the confirmation order, 'it cannot later complain about a certain provision contained in a confirmed plan, even if such a provision is inconsistent with the Code.'"[88] In accord with the Tenth Circuit's decision in *Andersen*, the *Pardee* court further concluded that "confirmation orders are final orders that are given preclusive effect."[89]

In emphasizing adherence to statutory restrictions, the Tenth Circuit retreated from its holding in *Andersen*, most notably with the decisions in *In re Poland*[90] and *In re Mersmann*.[91] Following the consolidation of two Chapter 13 bankruptcy cases on appeal, the Tenth Circuit in *Mersmann* overturned its decision in *Andersen*,[92] stating that a discharge-by-

---

86. Section 1328 of the Bankruptcy Code governs the discharge of debts as provided in reorganization plans and states, in pertinent part:

> (a) Subject to subsection (d), as soon as practicable after completion by the debtor of all payments under the plan, and in the case of a debtor who is required by a judicial or administrative order, or by statute, to pay a domestic support obligation, after such debtor certifies that all amounts payable under such order or such statute that are due on or before the date of the certification (including amounts due before the petition was filed, but only to the extent provided for by the plan) have been paid, unless the court approves a written waiver of discharge executed by the debtor after the order for relief under this chapter, the court shall grant the debtor a discharge of all debts provided for by the plan or disallowed under section 502 of this title, except any debt—
>
> . . .
>
> (2) of the kind specified in section 507(a)(8)(C) or in paragraph (1)(B), (1)(C), (2), (3), (4), (5), (8), or (9) of section 523(a) . . . .

11 U.S.C. § 1328(a)(2) (2006).

87. *In re Pardee*, 193 F.3d at 1086.

88. *Id.*

89. *Id.* at 1087.

90. Poland v. Educ. Credit Mgmt. Corp. (*In re* Poland), 382 F.3d 1185, 1189 (10th Cir. 2004) ("Because neither the plan nor the discharge order in this case contain any type of finding of undue hardship, we hold that *Andersen* does not apply and that the student loan debt is not discharged."). *Poland* is distinguished from *Andersen* in that the former case involved confirmation of a debt adjustment plan that did not mention undue hardship. *Id.* at 1188.

91. *In re* Mersmann, 505 F.3d 1033, 1038 (10th Cir. 2007) (overturning its decision in *Andersen* on the basis that confirmation of a plan which discharges a student loan debt absent an adversary proceeding violates the Bankruptcy Code and Rules).

92. *Id.*

declaration "is undeserving of res judicata effect" on the grounds that it does not conform to the Bankruptcy Code and Rules.[93] The *Mersmann* court further added that "[t]o give preclusive effect to a confirmation order based solely on § 1327(a) deprives the Bankruptcy Code and Rules of a coherent reading, fails to give full effect to all of their provisions, and undermines the clear will of Congress."[94]

C. Due Process Challenges

In *Mullane v. Central Hanover Bank & Trust Co.*, the Supreme Court stated that due process fundamentally requires "notice reasonably calculated, under all the circumstances, to apprise interested parties of the pendency of the action and afford them an opportunity to present their objections."[95] Commencement of an adversary proceeding to determine whether a student loan debt is dischargeable under the undue hardship exemption requires "highly specific service of process requirements" in accordance with Bankruptcy Rules 7001(6), 7003, and 7004.[96] Consequently, courts—particularly in the Fourth, Sixth, and Seventh Circuits—have rejected confirmed Chapter 13 plans on the basis that discharge-by-declaration provisions provide insufficient notice to creditors of a debtor's intention to discharge a student loan obligation.[97]

Both the *Andersen* and *Pardee* courts were conspicuously silent on the more substantive issue of due process that underlies the requirement for an adversary proceeding to determine student loan dischargeability.[98] The inference is that both Courts represented the view that sufficient notice is provided to the creditor merely by mailing the creditor a copy of the bankruptcy plan containing the offending provision and that due process in a bankruptcy proceeding to discharge student loan debt does demand an adversary proceeding.

In re Banks was the first principal case where the Court invoked due process concerns to justify its decision to affirm the district court's ruling against permitting the debtor to discharge by declaration the post-petition interest that had accrued on his student loans during Chapter 13 proceedings.[99] As the Fourth Circuit in *Banks* stressed, the creditor in this case was denied the "heightened degree of notice,"[100] as required by statute

93. *Id.* at 1048.

94. *Id.* at 1048–49.

95. Mullane v. Central Hanover Bank & Trust Co., 339 U.S. 306, 314 (1950).

96. *In re Mersmann*, 505 F.3d at 1043 (citing FED. R. BANKR. P. 7001(6), 7003, 7004).

97. *Id.* at 1046.

98. *See generally In re* Andersen, 179 F.3d 1253 (10th Cir. 1999) (not addressing concern of due process); *In re* Pardee, 193 F.3d 1083 (9th Cir. 1999) (not addressing concern of due process).

99. *In re* Banks, 299 F.3d 296, 302–03 (4th Cir. 2002).

100. *Id.* at 299.

under the circumstances: "The creditor received 'notice' of the plan provision pursuant to Bankruptcy Rule 2002 rather than service of process under Bankruptcy Rule 7004."[101] While disclaiming that its decision signified constitutional requirements for a complaint and summons in cases of discharging a student loan,[102] the Fourth Circuit concluded that due process entitled the student-loan creditor in *Banks* to all of the procedural protections specified under the Bankruptcy Code and Rules "before an order binding the party [could] be afforded preclusive effect."[103]

Adopting the Fourth Circuit's line of reasoning and holding in *Banks*, the Seventh Circuit in *In re Hanson* concluded that the failure of the student-loan debtor to serve the creditor-defendant with a summons and complaint for an adversary proceeding precluded the creditor "the opportunity of presenting an objection prior to the adjudication of its rights."[104] The Seventh Circuit held that the debtor's failure to comply with the Bankruptcy Code and Rules rendered the bankruptcy discharge order void under Rule 60(b)(4), because the creditor was not afforded the required higher level of due process notice.[105] The Seventh Circuit went a step further than the Fourth Circuit in *Banks*, however, and interjected legislative intent as well as justifications based on impropriety for its decision:

> The decision[] in *Banks* [has] greater persuasive force because [it is] consistent with Congress' unmistakable intent to make student loan debt nondischargeable absent a showing of undue hardship. Moreover, cases like *Andersen* and *Pardee* permit debtors to flout both substantive and procedural provisions of the Bankruptcy Code and Rules through a meaningless incantation of undue hardship in their proposed plans. Although we recognize the strong policy favoring finality of confirmation orders, due process entitles creditors to the heightened notice provided for by the Bankruptcy Code and Rules, and the dictates of due process trump policy arguments about finality.[106]

The Sixth Circuit in *In re Ruehle* joined with a growing number of courts in advancing the view that a student loan debt discharged by declaration in the absence of an undue hardship determination violates a creditor's substantial due process.[107] The *Ruehle* court echoed the Seventh

---

101. *Id.* at 301.

102. *Id.* at 303 n.4.

103. *Id.* at 302.

104. *In re* Hanson, 397 F.3d 482, 486 (7th Cir. 2005) ("We embrace the analysis and holding of the Fourth Circuit in *Banks* . . . .").

105. *Id.* at 487.

106. *Id.* at 486.

107. *In re* Ruehle, 412 F.3d 679, 685 (6th Cir. 2005) ("Due process is not to be sliced, diced and disguised with sauce. Due process must be served whole, without garnish.") (quoting *In re* Ruehle, 296 B.R. 146, 165 (Bankr. N.D. Ohio 2003))).

Circuit in *Hanson*, also noting reasons of legislative intent and statutory compliance for revoking discharge-by-declaration provisions.[108] Finally, the *Ruehle* court pointed out the possible ethical and moral implications of judicial decisions, like those made by the courts in *Andersen* and *Pardee*, that uphold student loan discharge by declaration on the basis of finality: such judicial behavior "enriches and emboldens those who take what is not theirs and legitimizes it with court sanction."[109]

III. THE BANKRUPTCY COURT'S ORDER CONFIRMING DEBTOR'S LOAN DISCHARGE PLAN IN THE ABSENCE OF AN UNDUE HARDSHIP FINDING AND ADVERSARY PROCEEDING IS NOT VOID UNDER RULE 60(B)(4)

In *United Student Aid Funds, Inc. v. Espinosa*, the United States Supreme Court held that an order issued by the bankruptcy court to confirm the discharge of a student loan debt in the absence of an undue hardship finding and adversary proceeding is not void under Rule 60(b)(4).[110] The Court based its decision on three principal determinations: (1) The bankruptcy court's judgment was not premised on a jurisdictional error;[111] (2) Espinosa's failure to serve United with a summons and complaint did not constitute a violation of the creditor's constitutional right to due process;[112] and (3) the bankruptcy court's lack of statutory authority does not fall within the narrow scope under which Rule 60(b)(4) relief may be granted.[113]

After providing an extensive précis of the factual background of this case, Justice Thomas, writing for a unanimous Court, began the analysis by examining the relevant statutory law pertaining to the dischargeability of student loan debt in bankruptcy.[114] In particular, the Court noted § 1328(a), which governs debt discharge, and § 523(a)(8), which establishes the undue hardship exemption to the general rule that student loans are nondischargeable.[115] The Court pointed out that, although debt discharge is less restricted under Chapter 13 than other bankruptcy chapters,[116] Chapter 13 strictly limits student loan discharges by imposing the requirement for an adversary proceeding which is commenced by service of process.[117]

---

108. *Id.* at 684.

109. *Id.*

110. United Student Aid Funds, Inc. v. Espinosa, 130 S. Ct. 1367, 1380 (2010).

111. *Id.* at 1377–78.

112. *Id.* at 1378.

113. *Id.* at 1379.

114. *Id.* at 1376.

115. *Id.*

116. *Id.* (citing 8 COLLIER ON BANKRUPTCY ¶ 1328.01 (Alan N. Resnick & Henry J. Sommer, et al. eds., 15th ed. rev. 2008).

117. *Id.*

The Court next turned to establishing the analytical framework by which the Court would assess the applicability of a Rule 60(b)(4) motion for relief to the Bankruptcy Court's confirmation order at issue in this case.[118] In an important footnote, the Court expressly limited its holding to "relief from judgment under Rule 60(b)(4)," stating the caveat that "[w]e express no view on the terms upon which other provisions of the Bankruptcy Rules may entitle a debtor or creditor to postjudgment relief."[119] In presenting its framework for analysis, the Court began with the proposition that the Bankruptcy Court's confirmation order adopting Espinosa's debt restructuring plan constituted a final judgment.[120] The Court noted, quoting from its earlier decision in *Travelers Indemnity Co. v. Bailey*,[121] that "[o]rdinarily, 'the finality of [a] Bankruptcy Court's orders following the conclusion of direct review' would 'stan[d] in the way of challenging [their] enforceability.'"[122] However, as explained by the Court, Rule 60(b) furnishes an "exception to finality," which permits the party of an adverse judgment to seek relief "under a limited set of circumstances."[123] Under Rule 60(b)(4), the court has the authority "to relieve a party from a final judgment if 'the judgment is void.'"[124]

The Court then proceeded to explicate the legal meaning of a "void judgment," characterizing it as "one so affected by a fundamental infirmity that the infirmity may be raised even after the judgment becomes final."[125] This is distinguished from a judgment that is merely erroneous, which is not void.[126] The Court underscored the necessity that "[t]he list of such infirmities [be] exceedingly short," stating that "otherwise, Rule 60(b)(4)'s exception to finality would swallow the rule."[127] In fleshing out its analytical framework, the Court enumerated the limited circumstances ("infirmities") under which an application of Rule 60(b)(4) is permissible: "Rule 60(b)(4) applies only in the rare instance where a judgment is premised either on a certain type of *jurisdictional error* or on a *violation of due process* that deprives a party of notice or the opportunity to be heard."[128]

---

118. *Id.* at 1376–77.

119. *Id.* at 1376 n.8.

120. *Id.* at 1376.

121. Travelers Indem. Co. v. Bailey, 129 S. Ct. 2195, 2198 (2009) (holding that collateral challenges to the enforceability of an injunction are generally precluded because of the finality accorded to the Bankruptcy Court's orders after direct review).

122. *Espinosa*, 130 S. Ct. at 1376 (quoting *Bailey*, 129 S. Ct. at 2198).

123. *Id.* (quoting Gonzalez v. Crosby, 545 U.S. 524, 29 (2005)).

124. *Id.* (citing FED. R. CIV. P. 60(b)(4)); *see supra* text accompanying note 16.

125. *Id.* at 1377 (citing RESTATEMENT (SECOND) OF JUDGMENTS § 22 (1980)).

126. *Id.* (citing Hoult v. Hoult, 57 F.3d 1, 6 (1st Cir. 1995)).

127. *Id.*

128. *Id.* (emphasis added).

Turning to an examination of jurisdictional error, the Court promptly dispensed with an "arguable basis"[129] inquiry into the circumstances by which a jurisdictional error could have occurred, because, as the Court pointed out, United did not contend that the "Bankruptcy Court's error was jurisdictional."[130] Even if United had assigned jurisdictional error to the Bankruptcy Court's order, the Court stated that such an argument would have failed regardless, because (1) § 523(a)(8)'s statutory requirement "is a precondition to obtaining a discharge order, not a limitation on the bankruptcy court's jurisdiction,"[131] and (2) "the requirement that a bankruptcy court make [an undue hardship] finding in an adversary proceeding derives from the Bankruptcy Rules, which are 'procedural. . .' . . . [and] 'not jurisdictional.'"[132] On a technical level, the Court, therefore, found neither § 523(a)(8) nor Rule 7001(6) to be jurisdictional in nature.[133]

The Court next addressed United's contention that the Bankruptcy Court's order confirming Espinosa's proposed debt adjustment plan violated United's due process rights, because Espinosa failed to serve the creditor with a summons and complaint as required by the Bankruptcy Code and Rules.[134] The Court flatly rejected this argument by the petitioner, stating that "Espinosa's failure to serve United with a summons and complaint deprived United of a right granted by a procedural rule."[135] The Court added that United had the opportunity to challenge such a procedural defect by timely objecting to "this deprivation" or appealing the "adverse ruling."[136] Citing *Mullane v. Central Hanover Bank & Trust Co.*, the Court stressed that due process only "requires notice 'reasonably calculated . . . to apprise interested parties of the pendancy of action and afford them an opportunity to present their objections.'"[137] Due process, as stated by the Court, does not require "*actual* notice."[138] In light of this Court's precedent, the Court determined that "United [had] received *actual* notice of the filing and contents of Espinosa's plan," which, as the Court found, constituted

---

129. *Id.* (citing Nemaizer v. Baker, 793 F.2d 58, 65 (2d Cir. 1986)). "[A] court will be deemed to have plainly usurped jurisdiction only when there is a 'total want of jurisdiction' and no arguable basis on which it could have rested a finding that it had jurisdiction." *Nemaizer*, 793 F.2d at 65 (citing Lubben v. Selective Serv. Sys. Local Bd. No. 27, 453 F.2d 645, 649 (1st Cir. 1972)).

130. *Espinosa*, 130 S. Ct. at 1377.

131. *Id.* at 1377–78.

132. *Id.* at 1378 (citing Kontrick v. Ryan, 540 U.S. 443, 454 (2004)) (internal citation omitted).

133. *Id.*

134. *Id.*

135. *Id.*

136. *Id.*

137. *Id.* (citing Mullane v. Central Hanover Bank & Trust Co., 339 U.S. 306, 314 (1950)).

138. *Id.* (citing Jones v. Flowers, 547 U.S. 220, 225 (2006)) (emphasis added).

sufficient notice despite the lack of service of process.[139] In the absence of a clearly demonstrated jurisdictional error or due process violation, the Court concluded that the facts of this case did not support the finding that United was entitled to relief under Rule 60(b)(4).[140]

Finally, the Court confronted United's assertion that the Bankruptcy Court's confirmation order is a nullity because it lacked the statutory authority to confirm Espinosa's plan for an adjustment of his student loan debt without first making an undue hardship finding.[141] Addressing the petitioner's argument that § 523(a)(8) "imposes a 'self-executing limitation on the effect of a discharge order' that renders the order legally unenforceable . . . if it is not satisfied,"[142] the Court stated that the "self-executing" nature of § 523(a)(8) mandates that the Bankruptcy Court make an undue hardship determination regardless of whether such a finding was requested by the affected creditor.[143] As the Court clarified, the fact that § 523(a)(8) is self-executing does not mean that a hardship finding is a condition precedent for a valid and enforceable confirmation order.[144] Instead, the Court found that "[g]iven the Code's clear and self-executing requirement for an undue hardship determination, the Bankruptcy Court's failure to find undue hardship before confirming Espinosa's plan was a legal error."[145] The Court continued, stating that this legal error did not entitle United to relief under Rule 60(b)(4), because "United had notice of the error and failed to object or timely appeal."[146] In rebuking creditors that fail to adequately protect their interests, the Court pointedly declared that "Rule 60(b)(4) does not provide a license for litigants to sleep on their rights."[147]

IV. CONSEQUENCES OF UNITED STUDENT AID FUNDS, INC. V. ESPINOSA

The United States Supreme Court's decision in *United Student Aid Funds, Inc. v. Espinosa* reflects the Court's unanimous endorsement of the Ninth Circuit's position in *In re Pardee*: A student-loan creditor forfeits its right to collaterally attack a confirmation order as void if, after receiving adequate notice of the debtor's proposed plan containing the objectionable discharge-by-declaration provision, the creditor fails to object to the plan or timely appeal the Bankruptcy Court's judgment.[148] Contrary to the Ninth

139. *Id.* (emphasis in original).
140. *Id.*
141. *Id.*
142. *Id.*
143. *Id.* at 1379.
144. *Id.*
145. *Id.* at 1380.
146. *Id.*
147. *Id.*
148. *See id.*; *In re* Pardee, 193 F.3d 1083, 1086 (9th Cir. 1999).

Circuit's position, however, the Supreme Court interpreted § 1325(a) as imposing an affirmative obligation on Bankruptcy Courts to ensure that proposed plans conform to the statutory requirements prior to confirmation.[149] Nevertheless, the Supreme Court's decision to uphold the binding effect of confirmed plans that contain illegal discharge-by-declaration provisions flies in the face of the prevailing authority in circuit case court law.[150] Additionally, the *Espinosa* Court's decision, to a certain extent, places the onus of creditor protection against impermissible student loan discharges on the creditors themselves, who simply cannot rely on the Bankruptcy Court to meticulously scour proposed debt reorganization plans for illegal discharge-by-declaration provisions.

Espinosa stands for the proposition that constitutionally guaranteed due process does not demand the heightened level of notice provided by an adversary proceeding.[151] In examining both the "notice" and "opportunity" aspects of sufficient due process, the Court stated that knowledge alone of the filing and contents of a proposed debt reorganization plan provides adequate notice to creditors of impending litigation that might affect their interests.[152] In the Court's view, creditors have sufficient opportunity to state their side of the case by raising timely objections or engaging in the appeal process.[153] How, in the absence of an adversary proceeding, which involves civil litigation procedure, is sufficient due process achieved in a meaningful sense through the opportunity of raising objections during a confirmation hearing?[154] One Bankruptcy Court has argued that "due process is not satisfied by discharging [a] debt through a plan provision," and discharging a nondischargeable student loan obligation without an adversary proceeding subverts the due process safeguards imposed by the Bankruptcy Code and Rules.[155] Commenting on the deficiency of the confirmation process as an opportunity to be heard, the Fifth Circuit has observed that "[u]nlike an objection to a proof of claim, the filing of a plan does not generally initiate a contested matter with respect to a particular claim," because the plan is not "a vehicle through which objections are made."[156] In *Espinosa*, clearly the Court, adhering to precedent established in *Mullane*, refused to elevate the rather low standard of notice "reasonably calculated" to inform. Certainly, the mailing of the proposed plan should have placed the creditor in this case on reasonable notice that Espinosa intended to default on at least a portion of his student loan obligations;

149. *Espinosa*, 130 S. Ct. at 1381.

150. *See, e.g.*, *In re* Mersmann, 505 F.3d 1033, 1053 (10th Cir. 2007); *In re* Whelton, 432 F.3d 150, 155 (2d Cir. 2005).

151. *Espinosa*, 130 S. Ct. at 1378.

152. *Id.*

153. *Id.* at 1380.

154. Driscoll, *supra* note 29, at 1328.

155. *In re* Galey, 230 B.R. 898, 899 (Bankr. S.D. Ga. 1999).

156. Internal Rev. Serv. v. Taylor (*In re* Taylor), 132 F.3d 256, 261 (5th Cir. 1998).

however, the Court failed to adequately consider whether reasonable "opportunity" to be heard—the second, and equally important, prong of due process—is afforded within the specialized context of a confirmation hearing for a bankruptcy plan.

While the Court did not condone the practice of discharge by declaration,[157] its decision in *Espinosa* fails to resolve the inherent contradiction of a court order that sanctions a plan in violation of § 523(a)(8)'s express directive. The Court acknowledged the potential for the continuation of such "bad-faith litigation tactics" but refused to "expand[] the availability of relief under Rule 60(b)(4)," calling the mechanism an "[in]appropriate prophylaxis."[158] Instead, *Espinosa* stands for a very limited proposition[159]—one that effectively precludes creditors from challenging erroneously confirmed bankruptcy plans on the basis of Rule 60(b)(4)—and as such, will likely not have a broad practical impact.[160] The Court correctly entrusts the matter of deterring "improper conduct in bankruptcy proceedings" to Congress through, perhaps, the enactment of additional statutory provisions or amendments to existing ones.[161] The Court's opinion instructs the Bankruptcy Court to "make an independent determination of undue hardship before a plan is confirmed, even if the creditor fails to object or appear in the adversary proceeding."[162] Whether this directive by the Court will have any meaningful effect on the number of nondischargeable student loan debts that are discharged remains to be seen.

V. CONCLUSION

The Supreme Court's decision in *United Student Aid Funds, Inc. v. Espinosa* represents the Court's whole-hearted affirmation of the Ninth Circuit's position with respect to the binding effect of discharge-by-declaration provisions, thus ensuring the continued practice of plan proponents attempting to overcome the presumption of student loan nondischargeability through carefully crafted language for the purpose of evading their student loan obligations. As a result, the Court's decision, partly due to its limited stand, does little to resolve the fundamental tension between discharge-by-declaration provisions and the statutory requirements for student loan dischargeability. In admonishing bankruptcy courts to carefully scrutinize plans for discharge-by-declaration language prior to confirmation and then diminishing creditor-protection options by declaring relief under Rule 60(b)(4) unavailable, the Court's decision sends mixed

157. *Espinosa*, 130 S. Ct. at 1380.

158. *Id.* at 1382.

159. *Id.* at 1376 n.8.

160. George Klidonas, *Are Nondischargeability Provisions Jurisdictional?*, 22 AM. BANKR. INST. J., June 2010, at 70 (2010).

161. *Espinosa*, 130 S. Ct. at 1382.

162. *Id.* at 1381.

signals concerning the propriety of such provisions. The Court's decision in *Espinosa* has only accentuated the dilemma of discharge by declaration.

DOROTHEA K. THOMPSON

TENNESSEE LAW REVIEW
Member of the National Conference of Law Reviews

Volume 78 Winter 2011 Number 2

CONTENTS

ARTICLES

SUNLIGHT'S GLARE: HOW OVERBROAD OPEN GOVERNMENT LAWS CHILL FREE SPEECH AND HAMPER EFFECTIVE DEMOCRACY
Steven J. Mulroy 309

THE MODERN HISTORY OF PROBABLE CAUSE
Wesley MacNeil Oliver 377

WHY CIVIL RECOURSE THEORY IS INCOMPLETE
Christopher J. Robinette 431

COMMENTS

IN THE WAKE OF *THERASENSE & NISUS CORP.*, HOW CAN PATENT ATTORNEYS DEFEND *THEMSELVES* AGAINST ALLEGATIONS OF INEQUITABLE CONDUCT?
Ian G. McFarland 487

THE RELIGIOUS LAND USE AND INSTITUTIONALIZED PERSONS ACT: A PERSPECTIVE ON THE UNREASONABLE LIMITATIONS PROVISION
Yusuf Z. Malik 531

CASE NOTES

CONSTITUTIONAL LAW—FIFTH AMENDMENT—HONING THE PROTECTIONS OFFERED BY *MIRANDA* THROUGH SUPREME COURT BRIGHT-LINE RULEMAKING
Caitlyn Luedtke Elam 569

CONSTITUTIONAL LAW—SIXTH AMENDMENT RIGHT TO COUNSEL—FORFEITURE OF THE RIGHT TO COUNSEL IN TENNESSEE
Scott M. McLeod 589

CIVIL PROCEDURE—INTERVENTION—NONSOVEREIGN ENTITIES IN EQUITABLE APPORTIONMENT ACTIONS INVOKING ORIGINAL JURISDICTION
Chelsey J. Hadfield 613

A TRIBUTE TO BOB LLOYD

SEMPER FIDELIS
Doug Blaze 637

THREE CHEERS FOR BOB LLOYD
Brannon P. Denning 639

PILOT LLOYD
Thomas C. Galligan, Jr. 643

WHAT ABOUT BOB?
Joan MacLeod Heminway 649

SUNLIGHT'S GLARE: HOW OVERBROAD OPEN GOVERNMENT LAWS CHILL FREE SPEECH AND HAMPER EFFECTIVE DEMOCRACY

STEVEN J. MULROY*

I. INTRODUCTION 310
II. BACKGROUND 315
A. *State Open Meetings Laws* 315
1. Scope 318
2. Exceptions 322
3. Remedies 324
B. *Federal Open Meetings Law* 325
III. FIRST AMENDMENT DISCUSSION 326
A. *Generally* 326
B. *Content-Based or Content-Neutral?* 332
1. Applying Strict Scrutiny 339
2. Applying Intermediate Scrutiny 342
C. *Legislators as Public Employees* 343
IV. EQUAL PROTECTION ISSUES 348
A. *Generally* 348
B. *Strict Scrutiny* 349
C. *Rational Basis* 353
V. POLICY DISCUSSION 355
A. *Policy Problems with the Broader Open Meetings Laws* 355
1. General: Applying "Transparency" Consistently 355
2. Whether the Broader Version is Truly Necessary to Fulfill Open Meetings Law Goals 358
3. The Costs of Overly Broad Open Meetings Laws 360
a. The Main Costs Raised by Commentators 360
b. Other Costs 364
B. *Model Open Meetings Law* 367
VI. CONCLUSION 370
VII. APPENDIX 371

* Associate Professor, Cecil C. Humphreys School of Law, University of Memphis. The author is also a part-time County Commissioner in Shelby County, Tennessee. He gratefully acknowledges advice from Professors George Cochran of the University of Mississippi School of Law, Rick Hasen of Loyola Law School in Los Angeles, and Gene Shapiro and Lee Harris from the University of Memphis, as well as assistance from Andrew Pate and Karol Landers in the preparation of this Article.

I. INTRODUCTION

Item: A County Commissioner forced to miss an upcoming vote emails colleagues his suggestions on how to proceed. The email is later read aloud at a publicly noticed open meeting.

Item: State legislators working on a lengthy, complex, and controversial law agree publicly while in session to a tentative compromise designed to resolve a bitter partisan dispute. Because there is no time to draft language there in the chamber, the leading members of the Democratic and Republican factions meet later to draft compromise language that will be presented at the next regular public legislative session.

Item: An alderperson attends a Sierra Club meeting where members are discussing a controversial local environmental issue, only to find a fellow alderperson making a presentation. During the question-and-answer period, both field questions from the audience about the best way to resolve the issue.

Responsible lawmaker action or subversions of the democratic process? Under some versions of state "open meetings" laws, there is a good chance that each of these scenarios is illegal. These "sunshine laws" forbid elected officials from conferring with each other about matters coming before them outside of a properly noticed public meeting. While the laws are designed to prevent back-room deals in smoke-filled rooms, their broad definitions of "meeting" and "deliberation" can potentially cause more severe problems.

Although the contours of the state laws vary widely, most apply to informal conversations, phone calls, or emails that contain any substantive discussion of government policy issues; some apply even if there are only two participants.[1] Many make no exceptions for personnel matters,[2] items

---

1. *See, e.g.*, COLO. REV. STAT. ANN. § 24-6-402(2)(a) (West 2007); TENN. CODE ANN. § 8-44-102(c) (Supp. 1998). *See generally* ANN TAYLOR SCHWING, OPEN MEETING LAWS 271–72 (2d ed. 2000) (describing methods employed in various states for determining the number of participants required to place a gathering under open meeting restrictions).

2. *See* ARIZ. REV. STAT. ANN. § 38-431(4) (2007); ARK. CODE ANN. § 10-3-305(a) (West 2010); COLO. REV. STAT. ANN. § 24-6-402(1)(b) (West 2007); CONN. GEN. STAT. ANN. §1-200(2) (West 2008); DEL. CODE ANN. tit. 29, § 10002(b) (West 2010); FLA. STAT. ANN. § 286.011 (West 2008); GA. CODE ANN. § 50-14-1(a)(2) (West 2010); HAW. REV. STAT. § 92-10 (West 2010); IDAHO CODE ANN. § 67-2341 (West 2007); 5 ILL. COMP. STAT. ANN. 120/1.02 (West 2010); IND. CODE ANN. § 5-14-1.5-2(a)(1) (West 2007); IOWA CODE ANN. § 21.2 (West 2007); KAN. STAT. ANN. § 75-4317a (West 2008); KY. REV. STAT. ANN. § 61.805(1) (West 2010); ME. REV. STAT. ANN. tit. 1, § 401 (2007); MICH. COMP. LAWS ANN. § 15.262(b) (West 2009); MISS. CODE ANN. § 25-41-1 (West 2010); NEB. REV. STAT. ANN. § 84-1409(2) (LexisNexis 2008); N.C. GEN. STAT. ANN. § 143-318.10(d) (West 2008); N.D. CENT. CODE § 44-04-17.1(8)(a),(b) (2007); OHIO REV. CODE ANN. § 121.22(B)(2) (West 2009); OKLA. STAT. tit. 25 § 304(2) (2010); OR. REV. STAT. ANN. § 192.630(5) (West 2010); 65 PA. CONS. STAT. ANN. § 703 (West 2007); S.C. CODE ANN. § 30-4-70(e) (2005); S.D. CODIFIED LAWS § 1-25-1 (2010); TENN. CODE ANN. § 8-44-102(b)(1)(A) (Supp. 1998); TEX. GOV'T CODE ANN. § 551.001(4)(A) (West 2009); VT. STAT. ANN. tit. 1, § 310(2) (West

threatening individual privacy,[3] financial negotiations, or other topics traditionally considered appropriate for private discussion.[4] Most of these laws punish violations with criminal or civil penalties.[5] For these reasons, these laws raise significant issues regarding the overbreadth and chilling effect on discussion of "core value" speech involving political matters.

Additionally, in over fifteen states, the open meetings provisions apply to local government bodies but not the state legislature, or the provisions are substantially more lenient as applied to the state legislature.[6] There is an obvious appeal for state legislators drafting these laws to exempt

---

2010); WASH. REV. CODE ANN. § 42.30.020(4) (West 2010); WIS. STAT. ANN. § 19.81(2) (West 2009); WYO. STAT. ANN. § 16-4-402(a)(iii) (West 2010).

3. *See* ALASKA STAT. ANN. § 44.62.310(a) (West 2009); LA. REV. STAT. ANN. § 42:6.2(A)(2) (2009); MONT. CODE ANN. § 2-3-202 (2009); NEV. REV. STAT. ANN. § 241.015(2)(a)(1) (2007); N.D. CENT. CODE § 44-04-17.1(8)(a) (2007); R.I. GEN. LAWS ANN. § 42-46-2(1) (West 2009); TENN. CODE ANN. § 8-44-102(b)(1)(A) (Supp. 1998); UTAH CODE ANN. § 52-4-103(4)(a) (West 2009).

4. *See* ALASKA STAT. ANN. § 44.62.310(a) (West 2009); ARK. CODE ANN. § 10-3-305(a) (West 2010); CAL. GOV'T CODE § 54950 (West 2009); COLO. REV. STAT. ANN. § 24-6-402(1)(b) (West 2007); CONN. GEN. STAT. ANN. §1-200(2) (West 2007); DEL. CODE ANN. tit. 29, § 10002(b) (West 2010); GA. CODE ANN. § 50-14-1(a)(2) (West 2010); HAW. REV. STAT. § 92-2(3) (West 2010); 5 ILL. COMP. STAT. 120/1.02 (West 2010); IOWA CODE ANN. § 21.2 (West 2007); KY. REV. STAT. ANN. § 61.805(1) (West 2010); LA. REV. STAT. ANN. § 42:6.2(A)(2) (2009); ME. REV. STAT. ANN. tit. 1, § 401 (2007); MICH. COMP. LAWS ANN. § 15.262(b) (West 2009); MONT. CODE ANN. § 2-3-202 (2009); NEB. REV. STAT. ANN. § 84-1409(2) (LexisNexis 2008); NEV. REV. STAT. ANN. § 241.015(2)(a)(1) (2007); N.H. REV. STAT. ANN. § 91-A:2(I) (2009); N.J. STAT. ANN. § 10:4-8(b) (West 2009); N.D. CENT. CODE § 44-04-17.1(8) (2007); OHIO REV. CODE ANN. §121.22(B)(2) (West 2009); OKLA. STAT. tit. 25, §§ 304(2) (2010); 65 PA. CONS. STAT. ANN. § 703 (West 2007); UTAH CODE ANN. § 52-4-103(4)(a) (West 2009); VT. STAT. ANN. tit. 1, § 310(2) (West 2007); W. VA. CODE ANN. § 6-9A-1 (West 2009); WYO. STAT. ANN. § 16-4-402(a)(iii) (West 2010).

5. *See, e.g.*, N.D. CENT. CODE § 44-04-21.2(1) (2007) (allowing a criminal fine of up to $1,000 in cases of willful violation).

6. *See* ALASKA STAT. § 44.62.310 (2009); ARK. CODE ANN. § 10-3-305(a) (West 2010); CONN. GEN. STAT. ANN. § 1-225(a) (West 2008); GA. CODE ANN. § 50-14-1(a)(1) (West 2010); HAW. REV. STAT. § 92-10 (West 2010; 5 ILL. COMP. STAT. ANN. 120/1.02 (West 2010) ; IND. CODE ANN. § 5-14-1.5-2(a)(1) (2007); KY. REV. STAT. ANN. § 61.805(1) (West 2010); LA. REV. STAT. ANN. § 42:6.2(A)(2) (2009); MISS. CODE ANN. § 25-41-1 (West 2010); N.M. STAT. ANN. § 10-15-2(A) (West 2009); OR. REV. STAT. ANN. § 192.630 (West 2010); S.C. CODE ANN. § 30-4-70(e) (2005); TENN. CODE ANN. § 8-44-102(a) (Supp. 1998); TEXAS GOV'T CODE ANN. §§ 551.003, 551.046 (West 2009); WASH. REV. CODE ANN. § 42.30.020(1)(a) (West 2010); WYO. STAT. ANN. § 16-4-402(a)(iii) (West 2010); Abood v. League of Women Voters of Alaska, 743 P.2d 333, 334 (Alaska 1987) (holding that the legislature could exempt itself from the open meetings law); Coggin v. Davey, 211 S.E.2d 708, 710 (Ga. 1975); Sarkes Tarzian, Inc. v. Legislature of the State of Nev., 765 P.2d 1142, 1144 (Nev. 1988) (holding that the legislature could make rules exempting it from the open meetings law in some cases); Mayhew v. Wilder, 46 S.W.3d 760, 769–70 (Tenn. Ct. App. 2001) (holding that the General Assembly does not fall within the definition of "governing body" applicable to the open meetings law).

themselves. But the disconnect between the freedom of speech afforded state legislators and the severe restrictions on local legislators raises a legitimate question of equal protection.

To date, these issues have received surprisingly little attention. A handful of state court cases have dismissed free speech challenges to open meetings laws without giving the issue much significant analysis.[7] Cases discussing equal protection challenges are hard to find.[8]

Scholarship on this issue has been light. It has focused mostly on the policy disadvantages of sunshine laws,[9] in some cases just at the federal

---

7. *See* Cole v. State, 673 P.2d 345, 350 (Colo. 1983) (finding the statute proper in light of the public's right to receive information); People *ex rel.* Difanis v. Barr, 397 N.E.2d 895, 899 (Ill. App. Ct. 1979) (ruling that the statute does not restrict the content of speech but merely requiring the speech to be public); St. Cloud Newspapers, Inc. v. Dist. 742 Cmty. Sch., 332 N.W.2d 1, 7 (Minn. 1983) (holding that the public's interest in hearing the content of government meetings outweighs government officers' rights to speak in closed sessions); Sandoval v. Bd. of Regents of Univ., 67 P.3d 902, 907 (Nev. 2003) (finding that the statute did not violate the First Amendment because officials' comments were not restricted, as long as they were scheduled); Smith v. Pa. Emps. Benefit Trust Fund, 894 A.2d 874, 880 n.4 (Pa. Commw. Ct. 2006) (dismissing a free speech challenge on the grounds that the statute was intended to promote discussion); Hays Cnty. Water Planning P'ship v. Hays Cnty., 41 S.W.3d 174, 182 (Tex. Ct. App. 2001) (holding that the statute restricted only the place and time of speech). *But see* McComas v. Bd. of Educ., 475 S.E.2d 280, 290–91 (W. Va. 1996) (examining the free speech issue more closely, upholding the law's application where the entire board physically met in secret, but establishing a multi-factor test to determine when a narrower application might violate free speech).

8. The relative lack of court challenges might not be so surprising after all. The persons most motivated to bring such challenges are elected officials. They are precisely those most vulnerable to the media criticism sure to follow from a public court challenge seeking the right to secret deliberations.

9. *See* Mark Fenster, *The Opacity of Transparency*, 91 IOWA L. REV. 885, 908–09 (2006) (reciting criticisms of open meetings laws based on the need for some private deliberations among decision-makers); Nicholas Johnson, *Open Meetings and Closed Minds: Another Road to the Mountaintop*, 53 DRAKE L. REV. 11, 22–24 (2004) (arguing that because not all public officials are experienced public speakers, some time to prepare collectively prior to public discussion should be allowed); Michael A. Lawrence, *Finding Shade From the "Government in the Sunshine Act": A Proposal to Permit Private Informal Background Discussions at the United States International Trade Commission*, 45 CATH. U. L. REV. 1, 10–12 (1995) (arguing for allowing private deliberations at the International Trade Commission); Joseph W. Little & Thomas Tompkins, *Open Government Laws: An Insider's View*, 53 N.C. L. REV. 451, 452 (1975); James T. O'Reilly & Gracia M. Berg, *Stealth Caused by Sunshine: How Sunshine Act Interpretation Results in Less Information for the Public About the Decision-Making Process of the International Trade Commission*, 36 HARV. INT'L L.J. 425, 458 (1995); Kathy Bradley, Note, *Do You Feel the Sunshine? Government in the Sunshine Act: Its Objectives, Goals, and Effect on the FCC and You*, 49 FED. COMM. L.J. 473, 481–85 (1997) (discussing a number of problems with the Sunshine Act, especially the erosion of collegiality between officials); Randolph J. May, *Taming the Sunshine Act; Too Much Exposure Inhibits Collegial Decision Making*, LEGAL TIMES, Feb. 5, 1996, at 24. *But see* Devon Helfmeyer, Note, *Do Public Officials Leave Their*

administrative level.[10] Discussion of possible constitutional challenges to such laws has not been extensive.[11]

A recent case has changed this. In *Rangra v. Brown*,[12] the Fifth Circuit held that Texas' open meetings law was a content-based restriction on speech subject to "strict scrutiny" constitutional review.[13] The court reversed a district court decision dismissing a free speech challenge and remanded to the district court for reconsideration under the exacting "strict scrutiny" standard.[14] The case has raised the potential invalidity of open meetings laws as a national issue.[15] The Fifth Circuit decided to re-hear the case en banc.[16] It ultimately dismissed the case as moot after the plaintiff elected official had left office.[17] The dismissal based on mootness came over a vigorous dissent from Judge Dennis, who noted that Rangra still faced a potential renewed prosecution under the open meetings law.[18] The case has also inspired some scholarly commentary.[19]

The controversy over the Texas Open Meetings Act is ongoing. Represented by the same lawyer in *Rangra,* a group of local elected officials from several localities have filed suit challenging the law on free speech grounds.[20] The case went to a bench trial at the end of 2010, and the

---

Constitutional Rights at the Ballot Box? A Commentary on the Texas Open Meetings Act, 15 TEX. J. C.L. & C.R. 205, 213–20 (2010) (discussing free speech issues involved with the Texas open meetings law raised by the case of *Rangra v. Brown*, 576 F.3d 531 (5th Cir. 2009)).

10. *See* Fenster, *supra* note 9, at 908–09; Lawrence, *supra* note 9, at 10–12; O'Reilly & Berg, *supra* note 9, at 458.

11. *See* Mandi Duncan, Comment, *The Texas Open Meetings Act: In Need of Modification or All Systems Go?*, 9 TEX. TECH ADMIN. L.J. 315, 317–22 (2008) (reviewing the Texas Open Meetings Act and discussing the district court's decision in *Rangra v. Brown*, 576 F.3d 531 (5th Cir. 2009)); Anthony B. Joyce, Note, *The Massachusetts Approach to the Intersection of Governmental Attorney-Client Privilege and Open Government Laws*, 42 SUFFOLK U. L. REV. 957, 968 (2009) (recognizing the existence of some constitutional debate); Kevin C. Riach, Case Note, *Epilogue to a Farce: Reestablishing the Power of Minnesota's Open Meeting Law*—Prior Lake American v. Mader, 33 WM. MITCHELL L. REV. 681, 682 (2007) ("[It may be] unfair and economically inefficient to resolve [the clash between public information and effective litigation] by construing public officials' use of attorney-client privilege more narrowly than private parties' use.").

12. 566 F.3d 515 (5th Cir. 2009).

13. *Id.* at 521.

14. *Id.* at 522.

15. *See* Chuck Lindell, *Advocates Fear Ruling Will Void Open Meetings Laws*, AUSTIN AMERICAN-STATESMAN, May 17, 2009, http://www.statesman.com/news/content/news/stories/local/05/17/0517speech.html.

16. Rangra v. Brown, 576 F.3d 531, 532 (5th Cir. 2009).

17. Rangra v. Brown, 584 F.3d 206, 207 (5th Cir. 2009) (en banc).

18. *Id.* at 207–11 (Dennis, J., dissenting).

19. *See* Helfmeyer, *supra* note 9, at 213.

20. *See* City of Alpine v. Abbott, 730 F. Supp. 2d 630, 631 (W.D. Tex. 2010).

district court rendered a decision late March 2011.[21] The district court rejected the free speech challenge,[22] in part for reasons distinguishing the Texas open meetings law from the statutes that are the focus of this article.[23] Another appeal to the Fifth Circuit is expected.[24]

The *Rangra* decision and its sequel raise a legitimate question about the significant free speech issues raised by at least the broadest of the open meetings laws. Particularly where the law (unlike the law at issue in *Rangra*) applies to substantive conversations between only two or three legislators, or where it allows no exceptions for private discussions of truly sensitive matters, a broad open meetings law can cause greater damage to democracy than the harm it is designed to prevent.

While legislators, courts, and commentators unqualifiedly laud "government in the sunshine,"[25] too much of anything, even sunshine, is not necessarily a good thing. The broadest of open meetings laws chill needed deliberation and collegiality, prevent compromise, and make unrealistic demands on busy part-time local legislators. They transfer power to unelected staff and lobbyists, encourage the violation of individual privacy, and force conscientious local legislators to become casual lawbreakers. While we have enjoyed five decades of increasing sunshine, it might be time for some shade.

This Article examines the constitutionality of open meetings laws. It draws on case law, objective public commentary, and the author's own experience as a local legislator dealing with one of the strictest open meetings regimes in the nation. Part II provides background on these "sunshine laws" nationally, their typical provisions, and their policy rationales. Part III discusses the potential success of a free speech challenge to such laws. It examines the possible standards of review and argues that under any of them, the most broad-reaching of sunshine law provisions likely fail to pass muster. Part IV assesses an equal protection challenge to laws that exempt the state legislature. It concludes that such a challenge's success may turn on whether rational basis or heightened review applies and examines arguments for the use of each standard. Part V discusses

---

21. *See* Asgeirsson v. Abbott, No. P-09-CV-59, 2011 WL 1157624 (W.D. Tex. Mar. 25, 2011).

22. *Id.* at *35–36.

23. *Id.* at *25–28, *30–31 (holding that the statute passed intermediate and strict scrutiny in part because it allowed private speech among less than a quorum of the public body, and because it provided exemptions for specified categories of speech like personnel matters). *See infra* Sections III.B.1 and III.B.2.

24. Email from plaintiffs' counsel Rod Ponton to author, April 4, 2011 (on file with author).

25. *See, e.g.*, 5 U.S.C. § 552b (2006); Office of the Governor v. Winner, 858 N.Y.S.2d 871, 874 (N.Y. App. Div. 2008) ("It is preferable that government operations be conducted in the sunshine of daylight."); Alison K. Hayden, *Two Cheers for the Illinois Freedom of Information Act*, 98 ILL. B.J. 82 (2010) ("'[S]unshine laws' are important tools for pulling back the curtain that often surrounds those in power.").

policy criticisms of open meetings laws, argues for a "scaling back" of their scope, and proposes a model open meetings law which balances the need for public access with the need for officials to be able to confer with one another to engage in responsible decision making.

II. BACKGROUND

A. State Open Meetings Laws

All fifty states have some form of open meetings laws.[26] Almost all of these open meeting laws require public notice and public access when deliberations are held or when public business is discussed by a governmental body.[27] The majority of these statutes apply to local

26. *See* OPEN GOVERNMENT GUIDE, http://www.rcfp.org/ogg/index.php (last visited Mar. 12, 2011).

27. *See, e.g.*, ALA. CODE § 36-25A-2(6) (2010); ARIZ. REV. STAT. ANN. § 38-431(4) (2007); COLO. REV. STAT. ANN. § 24-6-402(1)(b) (West 2007); CONN. GEN. STAT. ANN. § 1-200(2) (2008); DEL. CODE ANN. tit. 29, § 10002(b) (West 2010); GA. CODE ANN. § 50-14-1(a)(2) (West 2010); HAW. REV. STAT. § 92-2(3) (West 2010); IDAHO CODE ANN. § 67-2341(6) (West 2007); 5 ILL. COMP. STAT. ANN. § 120/1.02 (West 2010); IND. CODE § 5-14-1.5-2(a)(1) (West 2007); KAN. STAT. ANN. § 75-4317a(a) (West 2008); KY. REV. STAT. ANN. § 61.805(1) (West 2010); LA. REV. STAT. ANN. § 42:4.2(1) (2009); ME. REV. STAT. ANN. tit. 1, § 401 (2007); MD. CODE ANN., STATE GOV'T § 10-502(g) (West 2007); MICH. COMP. LAWS ANN. § 15.262(b) (West 2009); MISS. CODE ANN. § 25-41-1 (West 2010); MO. REV. STAT. § 610.010(5) (West 2007); MONT. CODE ANN. § 2-3-202 (2009); NEB. REV. STAT. § 84-1409(2) (LexisNexis 2008); NEV. REV. STAT. ANN. § 241.015(2)(a)(1) (2007); N.H. REV. STAT. ANN. § 91-A:2(I) (2009); N.J. STAT. ANN. § 10:4-8(b) (West 2009); N.M. STAT. ANN. § 10-15-1(B) (West 2007); N.C. GEN. STAT. ANN. § 143-318.10(d) (West 2008); N.D. CENT. CODE § 44-04-17.1(8) (2007); OR. REV. STAT. ANN. § 192.610(5) (West 2010); 65 PA. CONS. STAT. ANN. § 703 (West 2007); R.I. GEN. LAWS ANN. § 42-46-2(1) (West 2009); S.C. CODE ANN. § 30-4-20(d) (2009); TENN. CODE ANN. § 8-44-102(b)(1)(A) (2010); TEX. GOV'T CODE ANN. § 551.001(4)(A) (West 2009); UTAH CODE ANN. § 52-4-103(4)(a) (West 2009); VT. STAT. ANN. tit. 1, § 310(2) (West 2010); VA. CODE ANN. § 2.2-3701 (West 2009); WASH. REV. CODE ANN. § 42.30.02(3) (West 2010); WYO. STAT. ANN. § 16-4-402(a)(iii) (West 2010); Brookwood Area Homeowners Ass'n. v. Mun. of Anchorage, 702 P.2d 1317, 1323 (Alaska 1985) (holding that every step of the decision-making process of a governmental unit transacting public business is subject to Alaska Stat. § 44.62.310 (2009)); Ark. Gazette Co. v. Pickens, 552 S.W.2d 350, 353 (Ark. 1975) (declaring that all deliberations of a governing body must be held in public because the public is entitled to learn of actions taken by the governing body and the reasoning behind such actions under Arkansas's Freedom of Information Act: Ark. Stat. Ann. § 12-2801–12-2807 (1967)); Frazer v. Dixon Unified Sch. Dist., 22 Cal. Rptr. 2d 641, 651 (Cal. Ct. App. 1993) (stating that deliberative meetings fall under the California open meetings law: Cal. Gov't Code § 54950 (West 2009)); Wolfson v. State, 344 So.2d 611, 614 (Fla. Dist. Ct. App. 1977) (declaring that it was the intent of the government to subject all steps of the decision-making process to Florida's Sunshine Law: Fla. Stat. § 286.011 (2009)); St. Cloud Newspapers, Inc. v. Dist. 742 Cmty. Sch., 332 N.W.2d 1, 6 (Minn. 1983) (defining "meetings"' to include all discussions regarding matters which foreseeably would be subject to the board's final action and therefore subject to the

government bodies[28] and usually apply to any associated boards, commissions, and related bodies appointed by local government bodies to transact government business.[29] In at least twenty-eight states, the "sunshine law" also covers the state legislature.[30]

---

then in-force open meetings law: Minn. Stat. § 471.705 (2009)); Goodson Todman Enter. v. City of Kingston Common Council, 550 N.Y.S.2d 157, 158–59 (N.Y. App. Div. 1990) (declaring that not just voting sessions, but the entire decision-making process, is subject to N.Y. Pub. Off. LAW § 7 (McKinney 2009)); *In re* Appeal of the Order Declaring Annexation Dated June 28, 1978, 637 P.2d 1270, 1272 (Okla. Civ. App. 1981) (stating that all of the decision-making process is subject to the Oklahoma open meetings law: Okla. Stat. tit. 25, § 304 (1977)); Appalachian Power Co. v. Pub. Serv. Comm'n, 253 S.E.2d 377, 381 (W. Va. 1979) (clarifying meetings subject to the West Virginia open meetings law: W. Va. Code § 6-9A-1 (2009)); State *ex rel.* Newspapers, Inc. v. Showers, 398 N.W.2d 154, 156 (Wis. 1987) (clarifying which meetings fall under the open meetings law of Wisconsin: Wis. Stat. § 19.81 (2007)); S.D. Op. Att'y Gen. 89-08, *1-2 (1989) (stating that "meetings" includes when a majority of the body meets and discusses official business, thereby triggering the South Dakota open meetings law: S.D. Codified Laws § 1-25-1 (2009)).

28. *See* ALA. CODE § 36-25A-2(4) (2010); ALASKA STAT. ANN. § 44.62.310(h) (West 2009); ARIZ. REV. STAT. ANN. § 38-431(6) (2007); ARK. CODE ANN. § 12-280 (1967); CAL. GOV'T CODE § 54950 (West 2009); COLO. REV. STAT. ANN. § 24-6-402(1)(d) (West 2007); CONN. GEN. STAT. ANN. § 1-200(2) (West 2008); DEL. CODE ANN. tit. 29, § 10002(c) (West 2008); FLA. STAT. ANN. § 286.011(1) (West 2008); GA. CODE ANN. § 50-14-1(a)(2) (West 2010); HAW. REV. STAT. § 92-2(1) (West 2010); IDAHO CODE ANN. § 67-2341(5) (West 2007); 5 ILL. COMP. STAT. ANN. § 120/1.02 (West 2010); IND. CODE ANN. § 5-14-1.5-2(a), (b) (West 2007); IOWA CODE ANN. § 21.2 (West 2007); KAN. STAT. ANN. § 75-4317 (a) (West 2008); KY. REV. STAT. ANN. § 61.805(2) (West 2010); LA. REV. STAT. ANN. § 42:4.2(2) (2009); ME. REV. STAT. ANN. tit. 1, § 401 (2007); MD. CODE ANN., STATE GOV'T § 10-502(h) (West 2007); MICH. COMP. LAWS ANN. § 15.262(a) (West 2009); MISS. CODE ANN. § 25-41-1 (West 2010); MO. REV. STAT. § 610.010(4) (2007); MONT. CODE ANN. § 2-3-202 (2009); NEB. REV. STAT. ANN. § 82-1409(1) (LexisNexis 2008); NEV. REV. STAT. ANN. § 241.015(3) (2007); N.H. REV. STAT. ANN. § 91-A:1-a (2009); N.J. STAT. ANN. § 10:4-8(b) (West 2009); N.M. STAT. ANN. § 10-15-1(B) (West 2009); N.C. GEN. STAT. ANN. § 143-318.10(b) (West 2008); N.D. CENT. CODE § 44-04-17.1(12) (2007); N.Y. PUB. OFF. LAW § 102(2) (McKinney 2010); OHIO REV. CODE ANN. § 121.22(1) (West 2009); OKLA. STAT. tit. 25, § 304(1) (2010); OR. REV. STAT. ANN. § 192.610(3),(4) (West 2010); 65 PA. CONS. STAT. ANN. § 703 (West 2007); R.I. GEN. LAWS ANN. § 42-46-2(3) (West 2009); S.C. CODE ANN. § 30-4-20(a) (2009); S.D. CODIFIED LAWS § 1-25-1 (2009); TENN. CODE ANN. § 8-44-102(b)(1) (2007); TEX. GOV'T CODE ANN. § 551.001(3) (West 2007); UTAH CODE ANN. § 52-4-103(7) (West 2007); VT. STAT. ANN. tit. 1, § 310(3) (West 2007); VA. CODE ANN. § 2.2-3701 (2007); WASH. REV. CODE ANN. § 42.30.02(1) (West 2007); W. VA. CODE ANN, § 6-9A-1 (West 2009); WIS. STAT. ANN. § 19.81(2) (West 2009); WYO. STAT. ANN. § 16-4-402(a)(ii) (West 2010). For a useful and comprehensive compilation of various state approaches by topic, see the "Open Government Guide" at http://www.rcfp.org/ogg/index.php.

29. *See* ALA. CODE § 36-25A-2(4) (2007); ARIZ. REV. STAT. ANN. § 38-431(6) (2007); ARK. CODE ANN. § 12-280 (1967); CAL. GOV'T CODE § 54950 (West 2009); DEL. CODE ANN. tit. 29, § 10002(c) (West 2010); FLA. STAT. ANN. § 286.011(1) (West 2008); HAW. REV. STAT. § 92-2(1) (West 2010); IND. CODE ANN. § 5-15-1.5-2(a),(b) (West 2007); IOWA CODE ANN. § 21.2 (West 2007); KAN. STAT. ANN. § 75-4317(a) (West 2008); KY. REV. STAT. ANN.

States began to pass comprehensive open meetings laws in the 1950s.[31] By 1959, twenty states had such laws, and by the mid-1970s, every state had a statute that imposed open meeting requirements on a wide variety of government bodies.[32] Many of these laws were significantly strengthened after the Watergate scandal, which was viewed by many as proof of the need for more "sunshine" in government.[33]

The animating policy behind these laws is that government business should be conducted in public with adequate notice so that citizens can attend.[34] This openness is necessary in a democracy so that the electorate can be adequately informed of how decisions are made and have an opportunity to offer meaningful input.[35] To this end, open meeting laws

---

§ 61.805(2) (West 2010); LA. REV. STAT. ANN. § 42:4.2(2)(h); ME. REV. STAT. ANN. tit. 1, § 402(2) (2007); MD. CODE ANN., STATE GOV'T § 10-502(g) (West 2007); MICH. COMP. LAWS ANN. § 15.262(a) (West 2009); MO. REV. STAT. § 610.010(4) (2007); MONT. CODE ANN. § 2-3-202 (2009); NEB. REV. STAT. § 82-1409(1) (2006); NEV. REV. STAT. ANN. § 241.015(3) (2007); N.H. REV. STAT. ANN. § 91-A:1-a (2009); N.J. STAT. ANN. § 10:4-8(a) (West 2009); N.M. STAT. ANN. § 10-15-1(B) (West 2009); N.C. GEN. STAT. ANN. § 143-318.10(b) (West 2008); N.D. CENT. CODE § 44-04-17.1(12) (2007); OKLA. STAT. tit. 25, § 304(1) (2010); OR. REV. STAT. ANN.§ 192.610(3), (4) (West 2010); 65 PA. CONS. STAT. ANN. § 703 (West 2007); R.I. GEN. LAWS ANN. § 42-46-2(3) (West 2009); S.C. CODE ANN. § 30-4-20(a) (2006); S.D. CODIFIED LAWS § 1-25-1 (2010); TENN. CODE ANN. § 8-44-102(b)(1) (2007); TEX. GOV'T CODE ANN. § 551.001(3) (West 2009); UTAH CODE ANN. § 52-4-103(7) (West 2007); VT. STAT. ANN. tit. 1, § 310(3) (West 2010); VA. CODE ANN. § 2.2-3701 (2007); WASH. REV. CODE ANN. § 42.30.02(1) (West 2010); W. VA. CODE ANN. § 6-9A-1 (West 2009); WIS. STAT. ANN. § 19.81(2) (WEST 2009); WYO. STAT. ANN. § 16-4-402(a)(ii) (West 2010).

30. *See* ARIZ. REV. STAT. ANN. § 38-431(6) (2009); ARK. CODE ANN. § 10-3-305(a) (West 1987); COLO. REV. STAT. ANN. § 24-6-402(1)(d) (West 2008); DEL. CODE ANN. tit. 29, § 10002(c) (West 2009); KAN. STAT. ANN. § 75-4318 (West 2008); KY. REV. STAT. ANN. § 61.805(2) (West 2010); ME. REV. STAT. ANN. tit. 1, § 402(a) (2007); MD. CODE ANN., STATE GOV'T § 10-502(h) (West 2009); MICH. COMP. LAWS ANN. § 15.262(a) (West 2009); MO. REV. STAT. § 610.010(4) (2009); NEB. REV. STAT. § 84-1409(1) (2009); N.J. STAT. ANN. § 10:4-8(a) (West 2009); N.M. STAT. ANN. § 10-5-2(A) (West 2008); N.C. GEN. STAT. ANN. § 143-318.10(b) (West 2009); N.D. CENT. CODE § 44-04-17.1(6) (2009); OHIO REV. CODE ANN. § 121.22(B)(1) (West 2009); OKLA. STAT. tit. 25, § 304.1 (2009); 65 PA. CONS. STAT. ANN. § 712 (West 2009); R.I. GEN. LAWS ANN. § 42-46-2(3) (West 2008); S.C. CODE ANN. § 30-4-20(a) (2008); TEX. GOV'T CODE ANN. § 551.001(3) (West 2009); UTAH CODE ANN. § 52-4-102(7) (West 2009); VA. CODE ANN. § 2.2-3701 (2009).

31. SCHWING, *supra* note 1, at 3.

32. *Id.*

33. *Id.*

34. *See, e.g.*, ARK. CODE ANN. § 25-19-102 (1996); ME. REV. STAT. tit. 1, § 4012 (1989); MD. CODE ANN. STATE GOV'T § 10-501(a) (LexisNexis 1995); NEB. REV. STAT. § 84-1408 (Supp. 1998); N.C. GEN. STAT. §143-318.9 (1996); TENN. CODE ANN. § 8-440101(a) (1993).

35. *See, e.g.*, ARK. CODE ANN. §25-19-102 (1996); CAL. GOV'T CODE § 54950 (West 1997); MISS. CODE ANN. § 25-41 (1991).

provide that deliberations concerning public business shall not occur in private conversations between members of a governing body.[36]

1. Scope

As a general matter, these laws are intended to be, and are, construed very broadly. Often, the statutes include provisions stating that they are to be interpreted broadly to effectuate the policy goals of openness and accountability.[37] Courts also regularly state that these acts are to be given a broad construction.[38] This liberal construction generally persists even where the statutes contain penal provisions,[39] although some states strictly construe the penal provision while broadly construing the rest of the statute.[40]

The broad scope of the acts is evident from the expansive definitions of "governing body" or a similar phrase. Even in states that itemize some entities for inclusion, the general definition is typically given a broad interpretation.[41] Most states employ a number of criteria, such as manner of creation or receipt of public funds, any or all of which may place a given entity under the open meetings restrictions.[42]

Generally, there is no requirement that official action be taken or that official communications be made for a gathering or communication among officials to be covered by the open meetings law and thus be forbidden unless part of a properly noticed public meeting. While some states have exceptions for meetings held merely for ministerial purposes such as fact-gathering,[43] or to clarify a previous decision,[44] the statutes, as a rule, reach

---

36. *See, e.g.*, ME. REV. STAT. tit. 1, § 401 (1989) ("It is further the intent of the Legislature that clandestine meetings, conferences or meetings held on private property without proper notice and ample opportunity for attendance by the public not be used to defeat the purposes of this subchapter.").

37. *See, e.g.*, DEL. CODE ANN. tit. 29, § 10001 (West 2010); IND. CODE ANN. § 5-14-1.5-1 (West 2007); IOWA CODE ANN. § 21.1 (West 2007).

38. *See, e.g.*, Parole & Prob. Comm'n v. Thomas, 364 So.2d 480, 481 (Fla. Dist. Ct. App. 1978); Wexford Cnty. Prosecuting Attorney v. Pranger, 268 N.W.2d 344, 348 (Mich. Ct. App. 1978); Grein v. Bd. of Educ., 343 N.W.2d 718, 723 (Neb. 1984).

39. *See, e.g.*, City of Fayetteville v. Edmark, 801 S.W.2d 275, 278 (Ark. 1990); State *ex rel* Murray v. Palmgren, 646 P.2d 1091, 1096–97 (Kan. 1982), *appeal dismissed*, 459 U.S. 1081 (1982).

40. *See, e.g.*, WIS. STAT. ANN. § 19.81 (West 2009); Kansas City Star Co. v. Shields, 771 S.W.2d 101, 104 (Mo. Ct. App. 1989). *But cf.* State v. Patton, 837 P.2d 483, 484 (Okla. Crim. App. 1992) (recognizing that the open meeting law is a penal statute and must be strictly construed).

41. SCHWING, *supra* note 1, at 51.

42. *Id.*

43. *See, e.g.*, IOWA CODE ANN. § 21.2(2) (West 2007) (defining "meeting" so as not to include ministerial or social gatherings wherein no policy is discussed); Holeski v. Lawrence, 621 N.E.2d 802, 805–06 (Ohio Ct. App. 1993) (holding that the Ohio open meetings law does not apply to fact-finding sessions).

broadly to cover any substantive discussion of relevant government action. For example, Colorado requires all "meetings" to be open and noticed, and defines "meeting" as any gathering "convened to discuss public business."[45] State courts have also taken a broad view of legislative intent in this area; for example, the Alaska Supreme Court has construed the open meetings law to reach "every step of the deliberative and decision-making process when a governmental unit meets to transact public business."[46]

One key element of any open meetings law is its definition of "meeting." Some states define it as an official gathering convened for the purpose of considering matters of public significance.[47] However, most states apply restrictions to any meeting, planned or unplanned, at which a group's members discuss its business.[48]

At least twenty-eight states qualify this by requiring that a quorum or majority of the public body be present at the meeting before placing the discussion under the statute.[49] Two states say that the law applies whenever a majority of a quorum is involved in the meeting or discussion.[50] Even where a quorum or "majority of a quorum" is the rule, officials may not

---

44. *See, e.g.*, Hillsboro-West End Neighborhood Ass'n v. Metro Bd. of Zoning Appeals, No. 01A01-9406-CH-00282, 1995 Tenn. App. LEXIS 120, at *11 (Tenn. Ct. App. Feb. 24, 1995) (holding no open meetings law violation where defendants met privately to clarify a prior zoning board ruling).

45. COLO. REV. STAT. ANN. § 24-6-402(1)(b) (West 2007).

46. Brookwood Area Homeowners Ass'n v. Anchorage, 702 P.2d 1317, 1323 (Alaska 1985).

47. *E.g.*, N.Y. PUB. OFF. LAW § 102(1) (McKinney 2010).

48. *See* Hinds Cnty. Bd. of Supervisors v. Common Cause of Miss., 551 So.2d 107, 122–23 (Miss. 1989) (allowing public officials to meet as long as no public business is discussed).

49. ALASKA STAT. ANN. § 44.62.310(h)(2) (West 2009); ARIZ. REV. STAT. ANN. § 38-431(4) (2007); CAL. GOV'T CODE § 54952.2(a), (b) (West 2010); DEL. CODE ANN. tit. 29, § 10002(b) (West 2009); GA. CODE ANN. § 50-14-1(a)(2) (West 1999); HAW. REV. STAT. § 92-2.5(a), (f) (West 1998); IND. CODE ANN. § 5-14-1.5-2(c) (West 2007); IOWA CODE ANN. § 21.2(2) (West 2010); KY. REV. STAT. ANN. § 61.810(1) (West 2005); LA. REV. STAT. ANN. § 42:4.2(A)(1) (2009); MD. CODE ANN., STATE GOV'T § 10-502(g) (West 2010); MICH. COMP. LAWS ANN. § 15.262(b) (West 2010); MONT. CODE ANN. § 2-3-202 (2007); NEV. REV. STAT. ANN. § 241.015(2) (2010); N.H. REV. STAT. ANN. § 91-A:2(I) (2009); N.J. STAT. ANN. § 10:4-8(b) (West 2009); N.M. STAT. ANN. § 10-15-2(B) (West 2010); N.C. GEN. STAT. ANN. § 143-318.10(d) (West 2010); N.D. CENT. CODE § 44-04-17.1(8)(a) (2007); OHIO REV. CODE ANN. § 121.22(B)(2) (West 2010); OKLA. STAT. ANN. tit. 25, § 304(2) (West 2010); OR. REV. STAT. § 192.610(5) (2009); 65 PA. CONS. STAT. ANN. § 274 (West 2010); S.C. CODE ANN. § 30-4-20(d) (2005); S.D. CODIFIED LAWS § 1-25-1 (2009); TEX. GOV'T CODE ANN. § 551.001(4) (Vernon 2007); UTAH CODE ANN. § 52-4-103(4) (West 2010); VT. STAT. ANN. TIT. 1, § 310(2) (2007); W. VA. CODE ANN. § 6-9A-2(4) (West 2010); WYO. STAT. ANN. § 16-4-402(a)(i), (iii) (2005).

50. 5 ILL. COMP. STAT. ANN. 120/1.02 (West 2010); KAN. STAT. ANN. § 75-4317a (West 2010).

circumvent the law's strictures by having a series of smaller meetings that cumulate to a quorum or majority of a quorum.[51]

At least one state statute expressly applies the notice requirement whenever two or more members discuss public business.[52] Several more states reach this result through interpretation of the statute.[53] Virginia's statute requires a minimum of three legislators for the law's requirement to be triggered.[54] In a few other instances, the statute does not reach communications among only two members, but such a communication has been interpreted as illegal when it was done with intent to violate the statute's provisions.[55]

Tennessee's open meetings law is an unusual case: it has been interpreted to be among the strictest in the nation, but that interpretation is very much subject to question. Its statute defines "meeting" as "the convening of a . . . body for which a quorum is required," and it explicitly excludes from this definition "a chance meeting of two or more members."[56] This would suggest that Tennessee adopts the quorum rule. However, the statute also states that "such chance meetings" shall not "be used to decide or deliberate public business in circumvention of the spirit" of the Open Meetings Act.[57] Seizing on this last bit of "loophole closer" language, an unpublished county court decision held that the Act applied to any substantive conversation between two or more members.[58]

But that is by no means clear from the statute. The "loophole closer" language could just as easily have been written to apply to situations where two or three members constituted a quorum, where serial meetings of two or three members were held by design to cumulate a quorum—a so-called "walking quorum"—or both. Prior Tennessee cases did not raise this question, either because they involved communications among a quorum[59]

---

51. *See, e.g.*, KAN. STAT. ANN. § 75-4318(f) (West 2009)

52. COLO. REV. STAT. ANN. § 24-6-402(2)(a) (West 2010).

53. *See* Bigelow v. Howze, 291 So.2d 645, 647 (Fla. Dist. Ct. App. 1974) (applying the statute to a conversation between two of five County Commissioners); McElroy v. Strickland, Knox Cnty. Ch., No. 168933-2, at *10 (Oct. 5, 2007); Ala. Op. Att'y Gen 232-39 (construing statute to reach communication between two legislators); R.I. Opp. Att'y Gen. 92-06-09 (same).

54. VA. CODE ANN. § 2.2-3701 (West 2010).

55. *See* Sciolino v. Ryan, 440 N.Y.S.2d 795 (1981) (applying statute to just two, as long as they are deliberately evading the open meetings law); Mayor & City Council v. El Dorado Broad Co., 544 S.W.2d 206, 208 (Ark. 1976) (requiring two or more, but only when the mayor or a council member calls the meeting for the purpose of discussing public business); Haw. Op. Att'y Gen. 85-27 (stating that law possibly applies to two members, if the meeting is deliberate).

56. TENN. CODE ANN. § 8-44-102(b)(2), (c) (2010).

57. *Id.* § 8-44-102(c).

58. *McElroy*, No. 168933-2, at *10.

59. *See* Johnston v. Metro. Gov't of Nashville and Davidson Cnty., 320 S.W.3d 299, 310-12 (2009) (finding a violation for a series of emails between the whole metropolitan

or found communications among less than a quorum not to violate the Act for independent reasons.[60] Some appellate court cases have suggested without deciding that violations would involve a quorum.[61] An Attorney General opinion noted that whether a quorum was required presented a "difficult question," lacking "any definitive answer;" it concluded by issuing, as "cautious advice," the suggestion that local legislators err on the side of caution by avoiding substantive discussion among two or more members.[62] Local legislators and their in-house counsel have proceeded accordingly ever since, with the prevailing view that communications among any two members can violate the Act.[63] Given the constitutional

---

council); Waste Mgmt., Inc. of Tenn. v. Solid Waste Region Bd. of Metro. Gov't, No. M2005-01197-COA-R3-CV, 2007 WL 1094131, at *2 n.9 (Tenn. Ct. App. April 11, 2007) (citing Bordeaux Beautiful, Inc. v. Metro. Gov't, Davidson Cnty. Ch., No. 04-1513-III (June 4, 2004) (failing to overturn a lower court decision regarding a meeting among a quorum)); Grace Fellowship Church of Loudon Cnty., Inc. v. Lenoir City Beer Bd., No. E2000-02777-COA-R3-CV, 2002 WL 88874 (Tenn. Ct. App. Jan. 23, 2002) (vacating the Beer Board's decision to grant a permit where a quorum was present at the meeting in question); Englewood Citizens for Alternate B v. Town of Englewood, No. 03A01-9803-CH-00098, 1999 WL 419710, at *6 (Tenn. Ct. App. June 24, 1999) (finding that an improperly-noticed meeting among more than a quorum of the town Board of Commissioners was a violation of the OMA); Abou-Sakher v. Humphreys Cnty., 955 S.W.2d 65, 69–70 (Tenn. Ct. App. 1997) (finding that a meeting among a quorum of the airport authority violated the OMA); State *ex rel.* Akin v. Town of Kingston Springs, No. 01-A-01-9209-CH00360, 1993 WL 339305, at *2 (Tenn. Ct. App. Sept. 8, 1993) (finding an OMA violation, later cured by public meetings; though the court did not specify the number present at the disputed "work sessions," it implied that all members attended); State *ex rel.* Matthews v. Shelby Cnty. Bd. of Comm'rs, 1990 WL 29276, at *8 (Tenn. Ct. App. March 21, 1990) (reversing a dismissal of an OMA complaint when a "walking quorum" decided the issue amongst themselves through serial, individual discussions); Sharondale Constr. Co. v. Metro. Knoxville Airport Auth., 1989 WL 109470, at *3 (Tenn. Ct. App. Sept. 22, 1989) (affirming a dismissal for failure to allege particularized facts leading to the conclusion that an observed conversation was a "meeting;" the appellate court mentioned that the number of attendees, specifically relative to a quorum, would be relevant to the issue).

60. *See* Metro. Air Research Testing Auth., Inc. v. Metro. Gov't of Nashville, 842 S.W.2d 611, 618–19 (Tenn. Ct. App. 1992) (finding no violation where a number of city officials less than a quorum met without notice with a purchasing agent who was not bound by their recommendations); Univ. of Tenn. Arboretum Soc., Inc. v. City of Oak Ridge, 1983 WL 825161 (Tenn. Ct. App. May 4, 1983) (finding that a meeting of less than a quorum did not violate the OMA where no official business was discussed).

61. *See* Roberson v. Copeland, No. 85-199-II, 1985 WL 3524, at *5 (Tenn. Ct. App. Nov. 5, 1985) (making special note of the presence of a quorum); Dorrier v. Dark, 537 S.W.2d 888, 893 (Tenn. 1976) (rejecting a vagueness challenge in part because "the existence or non-existence of a quorum and whether or not they are in the course of deliberation" would almost always be clear to members of public bodies).

62. Tenn. Op. Att'y Gen. No. 88-169 (Sept. 19, 1988).

63. Craig E. Willis, *Sunshine Law Update*, 45 TENN. BAR J. No. 6, at 6–7 (2009).

issues raised in this Article, and the rule that statutes be construed to avoid constitutional issues,[64] this prevailing view is open to serious question.

At any rate, it is this latter category of state open meeting laws—ones affecting communications between only two or three members—that is most troubling from a First Amendment perspective.

2. Exceptions

The open meetings laws typically extend to indirect communications: a written, telephonic, or electronic communication will not escape the restrictions on that basis alone.[65] Most states treat mail correspondence as posing no less risk of abuse than a clandestine meeting.[66] This same reasoning applies to electronic letters in the form of email or similar technologies.[67] Telephone conversations have also been an issue and have been the subject of similar rulings in many states.[68]

There are some exceptions to the laws allowing for "executive sessions" concerning matters best discussed in private.[69] An executive session is typically defined as "a session closed to the public."[70] Courts have recognized that legislators sometimes need to debate an issue free from the pressures of partisans or interest groups.[71] The executive session exception does not allow legislators to simply hold secret meetings and then retroactively justify them according to the criteria for executive sessions.[72] Each statute has a protocol for a motion to hold an executive session, and the body must pass such a motion before holding the closed meeting.[73] Discussion within the executive session must then be limited to the subject

---

64. *See, e.g.*, Jerman v. Carlisle, McNellie, Rini, Kramer, & Ulrich, LPA, 130 S. Ct. 1605, 1635 (2010); State v. Burkhart, 58 S.W.3d 694, 697–98 (Tenn. 2001).

65. SCHWING, *supra* note 1, at 291.

66. *See, e.g.*, City Council v. Cooper, 358 So.2d 440, 441 (Ala. 1978) (finding that continued operation of city government by mailed ballots would irreparably harm citizens in light of the open meetings law); Common Cause v. Sterling, 147 Cal. App. 3d Supp. 518, 524 (Cal. Ct. App. 1983) ("[A]greeing to advise the city council members not to take future action by means of circulated letter . . . did violate the [act].").

67. SCHWING, *supra* note 1, at 293.

68. *See, e.g.*, Bd. of Trustees v. Bd. of County Comm'rs, 606 P.2d 1069, 1073 (Mont. 1980) (holding that a telephone conversation counted as a "meeting" requiring conformity with the open meetings statute); Hitt v. Mabry, 687 S.W.2d 791, 795 (Tex. Ct. App. 1985) (applying the law to telephone conversations).

69. SCHWING, *supra* note 1, at 358.

70. Sanders v. City of Fort Smith, 473 S.W.2d 182, 183 (Ark. 1971). *See generally* SCHWING, *supra* note 1, at 360.

71. *See* Pulitzer Publ'g Co. v. McNeal, 575 S.W.2d 802, 805 (Mo. Ct. App. 1978); Common Cause of Utah v. Utah Pub. Serv. Comm'n, 598 P.2d 1312, 1313–14 (Utah 1979).

72. SCHWING, *supra* note 1, at 361.

73. *See, e.g.*, Caldwell v. Lambrou, 391 A.2d 590, 593 (N.J. Super. Ct. Law Div. 1978).

matter contemplated in the motion, even if further issues arise in the meeting that would also meet the criteria for an executive session.[74]

Legislative bodies may not close a session for just any reason. Closure must fit a prescribed subject matter exception. Some states allow closed discussion for pending or anticipated litigation with counsel,[75] personnel matters,[76] matters affecting an individual citizen's privacy,[77] discussion of trade secrets,[78] or other topics.[79] Similarly, some states define "meeting" so as not to include one or more of these designated sensitive topics, or otherwise permit the requisite number of legislators to discuss such a topic without triggering the open meetings law.[80] A full list of topical exceptions, by state, is set out in the Appendix.[81]

However, almost no state has accepted all these topics, and many states have few or none.[82] Many states admit no exception for personnel matters, for example.[83] Some would even require that ongoing financial negotiations between the local government and an outside entity be carried out in public.[84]

Tennessee is a good example. By its terms, the Tennessee Open Meetings Act exempts, from public notice requirements, only discussions of trade secrets or consultation with counsel regarding pending litigation.[85] The statute itself does not even provide the allowance for private consultation with counsel:[86] such an exception was mandated by Tennessee

---

74. *See, e.g.*, NEB. REV. STAT. § 84-1410(3) (2009).

75. *E.g.*, CAL. CONST. art. IV, § 7(c)(1)(C); N.Y. PUB. OFF. LAW § 105(1)(d) (McKinney 1988).

76. *See, e.g.*, CAL. CONST. art. IV, § 7(c)(1); MICH. COMP. LAWS ANN. § 15.268(a), (f) (West 2010).

77. *E.g.*, MONT. CODE ANN. § 2-3-203(3) (2009) ("[I]f and only if . . . the demands of individual privacy clearly exceed the merits of public disclosure."); VA. CODE ANN. § 2.2-3711(A)(4) (West 2010).

78. *See, e.g.*, S.C. CODE ANN. § 30-4-40(a)(1) (2010).

79. *See, e.g.*, MD. CODE ANN., STATE GOV'T § 10-508(a)(10) (West 2010); MISS. CODE ANN. § 25-41-3(a) (West 2010); WASH. REV. CODE ANN. § 42.30.110(1)(a) (West 2010).

80. *E.g.*, HAW. REV. STAT. § 92-2.5(c) (West 2009) (allowing any group of less than a quorum to discuss selection of board officers without limitation). *Contra* Caldwell v. Lambrou, 391 A.2d 590, 593 (N.J. Super. Ct. Law Div. 1978) (requiring passage of a formal resolution, in a public meeting, indicating that an exception applies before any private meeting can occur). Many states require a formal motion to hold an executive session in order for a body to invoke an exception to the open meetings law, but these states may allow a few exceptions to this rule.

81. I am grateful for Nathaniel Terrell's assistance in the preparation of this Appendix.

82. *E.g.*, N.D. CENT. CODE § 44-04-19 (2009); TENN. CODE ANN. § 8-44-102 (2010).

83. *E.g.*, NEV. REV. STAT. ANN. § 241.020 (2010); N.D. CENT. CODE § 44-04-19 (2009); TENN. CODE ANN. § 8-44-102 (2010); UTAH CODE ANN. § 52-4-201 (West 2010).

84. *E.g.*, MONT. CODE ANN. § 2-3-203(3) (2009); NEV. REV. STAT. ANN. § 241.020 (2010); N.D. CENT. CODE § 44-04-19 (2009).

85. TENN. CODE ANN. § 8-44-102.

86. *Id.*

courts.[87] In Tennessee, the open meetings law requirements are triggered whenever two or more members of the government body have a substantive discussion of any matter which is currently or about to be before them; even a meeting with an attorney must be open if the body engages in any decision making or deliberation.[88] Tennessee's law is one of the strictest in the nation.

3. Remedies

The remedies available under open meetings laws vary from state to state, but they generally involve suing for enforcement.[89] Standing for such lawsuits tends to be as broad as possible, with many statutes granting expanded standing for parties such as the news media.[90]

The statutes typically allow parties to obtain an injunction by showing a violation of the open meetings law.[91] Most states also provide for civil penalties,[92] many of which increase with multiple violations.[93] Although criminal penalties may be less attractive due to their higher standard of proof,[94] many statutes provide for criminal fines or even imprisonment.[95] Most importantly, many states have a mechanism for retroactively invalidating actions taken through an illegal deliberation.[96] A few states even provide a mechanism for removing violators from office.[97] Even where the statute explicitly provides for only injunctive relief, courts may

---

87. Van Hooser v. Warren Cnty. Bd. of Educ., 807 S.W.2d 230, 237 (Tenn. 1991).

88. TENN. CODE ANN. § 88-44-102.

89. SCHWING, *supra* note 1, at 471–72.

90. *E.g.*, Ark. Gazette Co. v. Pickens, 522 S.W.2d 350, 355 (Ark. 1975).

91. *E.g.*, GA. CODE ANN. § 50-14-5(a) (West 2009); TEX. GOV'T CODE ANN. § 551.142 (West 1994); WASH REV. CODE ANN. § 42.30.130 (West 2009).

92. *E.g.*, LA. REV. STAT. ANN. § 42:13 (2010), WIS. STAT. ANN. § 19.96 (West 2010).

93. *E.g.*, IDAHO CODE ANN. § 67-2347(4) (West 2010) (adding a $500 recidivism penalty for multiple violations within a twelve-month period).

94. SCHWING, *supra* note 1, at 509.

95. *E.g.*, ARK. CODE ANN. § 25-19-104 (West 2010) (making a violation a Class C misdemeanor, carrying the possibility of both fines and jail time); *see* Helfmeyer, *supra* note 9, at 227–30 (finding that at least nineteen state open meetings laws impose criminal penalties, with twelve of those including imprisonment as an option and the remaining seven providing for fines or removal from office only).

96. *E.g.*, ARIZ. REV. STAT. ANN. § 38-431.05(A) (2010) (requiring mandatory vacation of all decisions reached in illegal meetings); IOWA CODE ANN. § 21.6(3)(c) (West 2010) (allowing vacation of illegally-reached decisions at the court's discretion); Carter v. City of Nashua, 308 A.2d 847, 856 (N.H. 1973) (holding that the judiciary has discretion to vacate illegally-reached decisions absent an explicit statutory rule).

97. *E.g.*, ARIZ. REV. STAT. ANN. § 38-431.07(A) (2010) (granting courts discretion to remove officials who violate the open meetings law with intent to disenfranchise the public); HAW. REV. STAT. § 92-13 (West 2010) (granting courts discretion to summarily remove from office any individuals convicted of willful violations of the open meetings law).

retain equitable discretion to fashion additional relief, such as money damages.[98]

B. Federal Open Meetings Law

The federal analogue to the open meetings laws is the Sunshine Act, passed in 1976.[99] The Sunshine Act applies to federal agencies and requires that every "meeting" be held in a public forum pursuant to notice.[100]

However, the Sunshine Act is much narrower than its typical state counterpart. First, it applies only to federal agencies and not to the legislature.[101] As with many state legislatures, there is a natural temptation for those enacting a "sunshine" law to exempt themselves from its provisions. Second, a "meeting" is defined as an assembly of a quorum of the body.[102]

Further, there are no less than ten permissible exemptions allowed for a closed meeting. They involve discussions of (1) national defense; (2) personnel issues; (3) statutorily-protected information; (4) trade secrets; (5) accusations of criminal conduct or formal censure; (6) matters of personal privacy; (7) investigatory records; (8) information generated in the regulation of financial institutions; (9) information likely to produce financial speculation or threaten an institution's financial stability; and (10) information related to various legal proceedings.[103] The federal Freedom of Information Act has an almost identical list of exemptions applicable to requests for government records.[104]

Finally, the remedies provided under the federal version are weaker than state versions. There are no criminal or civil penalties. The court may *not* nullify a decision if it finds a violation. Aside from enjoining further violations and assessing court fees, the court may only order that the contents of the meeting be disclosed to the public.[105]

---

98. *See, e.g.*, TENN. CODE ANN. § 8-44-106(a) (2010) (empowering courts to impose penalties for violations); Forbes v. Wilson Cnty. Emergency Dist. 911 Bd., 966 S.W.2d 417, 421 (Tenn. 1998) (trial courts have equitable discretion to award monetary damages).

99. 5 U.S.C. § 552b (2006).

100. *Id.* § 552b(b).

101. *Id.* § 552b(a)(1).

102. *Id.* § 552b(a)(2).

103. *Id.* § 552b(c). To counterbalance this extensive list of exemptions, the Act provides for a presumption in favor of openness, allows a citizen to challenge a decision to close a meeting, and places the burden of proof in such a challenge on the agency. *Id.* § 552b(h)(1).

104. *See id.* § 552b(b) (listing exemptions analogous to all but items (5), (9) and (10), and adding exemptions for (i) geological information concerning wells and (ii) inter-agency or intra-agency memoranda); *see also id.* § 552b(c) (adding separate exemption, similar to (10) above, concerning certain information relevant to pending criminal investigations).

105. *Id.* § 552b(i).

III. FIRST AMENDMENT DISCUSSION

A. Generally

The open meetings statutes are relatively new in United States history and generally do not have common law antecedents.[106] States typically do not recognize a common law right to attend meetings of governmental bodies.[107] Further, courts do not recognize a constitutional right to have all meetings of public bodies be open to the public.[108]

Nor has the rule of open meetings long been part of our historical practice. The delegates at the Constitutional Convention in 1787 deliberated in secret,[109] as did the members of the first Congress who debated the Bill of Rights.[110] Congress began to open at least some of its meetings to the public early on, but congressional committee meetings have only been routinely opened to the public since 1970.[111] Even today, while congressional debates and committee meetings are open to the public, there is no legal restriction on members of Congress conferring in private to hold substantive discussions on public business. Indeed, the practice is quite frequent.

At first glance, it may seem that First Amendment concerns would weigh toward strict enforcement of open meetings laws. The United States Supreme Court has ruled that the freedom to speak includes the freedom to receive information.[112] Courts have indicated that the First Amendment grants the public some sort of right of access to certain government proceedings. For the most part, the cases have involved access to criminal proceedings and have provided a qualified right of access subject to limitations set by the trial judge.[113] Some lower courts have extended this

---

106. *See* SCHWING, *supra* note 1, at 1.

107. *See id.* (citing various state court cases).

108. *See, e.g.*, Minn. State Bd. v. Knight, 465 U.S. 271, 284 (1984) (recognizing that public bodies may constitutionally hold non-public sessions to transact business); Madison Joint Sch. Dist. v. Wis. Emp't Relations Comm'n, 429 U.S. 167, 175 n.8 (1976); Bi-Metallic Inv. Co. v. State Bd. of Equalization, 239 U.S. 441, 445 (1915) ("The Constitution does not require all public acts to be done in town meeting or an assembly of the whole.").

109. *See* United States v. Nixon, 418 U.S. 683, 705 n.15 (1974) (citing 1 M. FARRAND, THE RECORDS OF THE FEDERAL CONVENTION OF 1787 xi–xxv (1911)).

110. *See* GEORGE LANKEVICH, ROOTS OF THE REPUBLIC: THE FIRST HOUSE OF REPRESENTATIVES AND THE BILL OF RIGHTS 13, 22–23 (Gary D. Hermalyn, C. Edward Quinn & Lloyd Ultan eds., Grolier Educational 1996).

111. SCHWING, *supra* note 1, at 2. Though a few states had laws opening up isolated government bodies in the 1800s, the first comprehensive open meetings law did not pass until 1915. *Id.*

112. Va. State Bd. of Pharmacy v. Va. Citizens Consumer Council, Inc., 425 U.S. 748, 757 (1976).

113. *See generally* Press Enter. Co. v. Super. Ct. of Cal. (*Press Enterprise II*), 478 U.S. 1 (1986) (criminal preliminary hearing); Press Enter. Co. v. Superior Court of Cal. (*Press*

First Amendment analysis to civil court proceedings as well.[114] However, courts have not found a constitutional right of public access to legislative proceedings.[115] Indeed, in broad language regarding other types of nonjudicial government bodies, the Supreme Court has suggested the opposite.[116] At any rate, the question of public access to legislative meetings has been settled by the adoption of open meetings laws in all states.[117]

Even if the federal Constitution does not require the kind of right of public access guaranteed by these statutes, it is arguable that some of the protections afforded by these statutes may be required by particular state constitutions, which are free to provide greater individual liberty protection than the federal Constitution.[118] A few states have interpreted their state constitutions explicitly to guarantee public access to, or public notice of, the deliberations of public bodies,[119] but even they are subject to some limits.[120]

Enterprise I), 464 U.S. 501 (1984) (criminal jury selection); Globe Newspaper Co. v. Superior Court, 457 U.S. 596 (1982) (criminal trials); Richmond Newspapers, Inc. v. Virginia, 448 U.S. 555 (1980) (criminal trials); *In re* Washington Post Co., 807 F.2d 383 (4th Cir. 1986) (sentencing hearings).

114. *See, e.g.*, Grove Fresh Distribs., Inc. v. Everfresh Juice Co., 24 F.3d 893, 897 (7th Cir. 1994) ("Though its original inception was in the realm of criminal proceedings, the right of access [to judicial proceedings] has since been extended to civil proceedings because the contribution of publicity is just as important there. . . . [T]he right of access belonging to the press and the general public also has a First Amendment basis."); Publicker Indus. v. Cohen, 733 F.2d 1059 (3d Cir. 1984) (preliminary injunction hearing); *In re* Cont'l Ill. Sec. Litig., 732 F.2d 1302 (7th Cir. 1984) (hearing on motion to dismiss); *In re* Iowa Freedom of Info. Council, 724 F.2d 658 (8th Cir. 1984) (contempt hearing); Brown & Williamson Tobacco Corp. v. Fed. Trade Comm'r, 710 F.2d 1165 (6th Cir. 1983) (vacating the district court's sealing of documents filed in a civil action based on common law and First Amendment right of access to judicial proceedings); Newman v. Graddick, 696 F.2d 796 (11th Cir. 1983) (pre- and post-trial hearings); Doe v. Santa Fe Indep. Sch. Dist., 933 F. Supp. 647, 648–50 (S.D. Tex. 1996) (concluding that the right of the public to attend civil trials is grounded in the First Amendment as well as the common law).

115. *See, e.g.*, Minn. State Bd. v. Knight, 465 U.S. 271, 284 (1984) (stating that public bodies may constitutionally hold non-public sessions to transact business); Madison Joint Sch. Dist. v. Wis. Emp't Relations Comm'n, 429 U.S. 167, 175 n.8 (1976) (same); Bi-Metallic Inv. Co. v. State Bd. of Equalization, 239 U.S. 441, 445 (1915) ("The Constitution does not require all public acts to be done in town meeting or an assembly of the whole."); Flesh v. Bd. of Trs., 786 P.2d 4, 10 (Mont. 1990) (concluding that the closure of grievance hearing on privacy grounds did not violate the First Amendment).

116. *See, e.g.*, *Knight*, 465 U.S. at 284; *Wis. Emp't Relations Comm'n*, 429 U.S. at 175 n.8; *Bi-Metallic Inv. Co.*, 239 U.S. at 445; *Flesh*, 786 P.2d at 10.

117. SCHWING, *supra* note 1, at 2.

118. *Id.*

119. *See, e.g.*, Law & Info. Servs., Inc. v. Riviera Beach, 670 So.2d 1014 (Fla. Dist. Ct. App. 1996); Hayes v. Jackson Parish Sch. Bd., 603 So.2d 274 (La. Ct. App. 1992); Associated Press v. Bd. of Pub. Educ., 804 P.2d 376 (Mont. 1991).

120. *E.g.*, Eastwold v. New Orleans, 374 So.2d 172, 173 (La. Ct. App. 1979) (holding that meetings can be scheduled during normal business hours, even if this interferes with the

Indirect support for such access might also be found from state constitutional provisions on free speech, free press, the right of assembly, the right to petition for redress of grievances, and so forth.[121] However, there is little case law supporting such a reading of these state constitutional provisions.[122] Further, any such requirements would be trumped if found to be inconsistent with the federal Constitution.[123]

Insufficient attention has been given to the negative free speech implications of these laws. Clearly, they cause a substantial restriction on political speech.

No state court adjudicating a free speech challenge to its state's open meetings law has overturned the law on free speech grounds. Some state courts have stated in dicta that such laws in general raise significant free speech issues, though none have referenced the specific statute before them.[124] For example, the Florida Supreme Court has stated that the statute would violate the First Amendment if its law were construed "to prohibit any discussion whatever by public officials between meetings."[125] However, that court also suggested that a conventional interpretation barring substantive discussion of matters before the government body would likely pass muster.[126] Similarly, in *Dorrier v. Dark*, the Tennessee Supreme Court rejected such a free speech challenge on grounds peculiar to the Tennessee open meetings law: because there was no penalty other than invalidating the decision taken by the public body, the court reasoned, there was no significant "chilling effect" on free speech.[127] The court also noted that a free speech violation would likely lie if the law had criminal penalties, as many state open meetings laws do.[128]

---

ability of some individuals to attend).

121. *E.g.*, Maurice River Twp. Bd. of Educ. v. Maurice River Twp. Teachers' Ass'n, 475 A.2d 59 (N.J. Super Ct. App. Div. 1984) (holding that the constitutional rights of freedom of assembly and petition for redress of grievances create a right to access public meetings).

122. SCHWING, *supra* note 1, at 16.

123. *See* Contemporary Indus. Corp. v. Frost, 564 F.3d 981, 988 (8th Cir. 2009) (discussing the Supremacy Clause).

124. City of Miami Beach v. Berms, 245 So.2d 38, 41 (Fla. 1971); *see also* Dorrier v. Dark, 537 S.W.2d 888, 892 (Tenn. 1976) (explaining that a chilling effect on free speech would arise if the Tennessee statute, like most other open meeting statutes, punished violations with fines, criminal punishments, or removal from office).

125. *Berms*, 245 So.2d at 41.

126. *Id.* In addition, one state supreme court decision struck down the criminal provision of an open meetings law on vagueness grounds but did not reach the free speech issue. *See* Knight v. Iowa Dist. Ct. of Story Cnty., 269 N.W.2d 430, 432–34 (Iowa 1978) (finding the criminal provision vague because it did not specify what level of participation in an illegal meeting constituted illegal conduct).

127. *Dorrier*, 537 S.W.2d at 892.

128. *Id.*

More significant is the free speech discussion by the West Virginia Supreme Court of Appeals in *McComas v. Board of Education of Fayette County*.[129] In that case, the court upheld an application of the state's open meetings law where all but one of a board of education's members physically met in secret with the school superintendent to discuss business coming before it publicly the next day.[130] In that instance, of course, a quorum of the membership had met.[131]

The court instructively considered the kinds of meetings, gatherings, and informal conversations that might be covered under its sunshine law.[132] It stated that an interpretation that "precludes any off-the-record discussion between board members about board business would be both undesirable and unworkable—and possibly unconstitutional."[133] Such a "sweeping restriction" on public officials' ability to discuss "public issues in a private manner" would raise "serious questions" under the First Amendment.[134]

To avoid this constitutional issue, the Court adopted a more flexible, "common sense approach" which focused on the question of whether allowing a private conversation among officials under particular circumstances would "undermine the [sunshine law's] fundamental purposes."[135] Making this determination in turn requires consideration of many factors, none exhaustive or controlling: the content of the discussion; the number of members of the public body participating; the percentage of the public body this number represents; the identity of the absent members; the intentions of the members; the amount of planning involved; the duration of the conversation; the setting; and the possible effect on decision making.[136] As in this Article, the *McComas* court drew a distinction between conversations between two members of a body and conversations among a quorum of a body.[137] Explaining that "[n]umbers are relevant," the court emphasized the "difference between two members of a twenty-member public body having a conversation and fifteen of them having a cabal."[138]

McComas is unique among state court decisions in its detailed, nuanced approach to the free speech issues.[139] The *McComas* court recognized the

129. 475 S.E.2d 280 (W. Va. 1996).

130. *Id.* at 293.

131. *Id.* at 298–99.

132. *Id.* at 290.

133. *Id.*

134. *Id.* at 290 n.18.

135. *Id.* at 290.

136. *Id.* Under the facts in *McComas*, the Court held that the sunshine law was appropriately applied where an actual physical meeting was planned and attended by four-fifths of a school board's members with the intent to discuss information relevant to an issue coming before the board. *Id.* at 293.

137. *Id.* at 291.

138. *Id.*

139. *See id.* at 280.

legitimate state interest in ensuring that the public have a "meaningful opportunity to respond to, or hold officials accountable for, their private deliberations."[140] However, the court also rejected as overbroad any restrictions on private conversations among elected officials where such restrictions are not actually required to further those governmental interests.[141] Although the court did not say so explicitly, its approach was not unlike one requiring that the open meetings law be "narrowly tailored" to further the state's compelling governmental interests.

Five other state court cases have upheld open meetings laws against free speech challenges.[142] Crucially, each of those cases involved physical meetings among a quorum or more of the members.[143] As explained below, since a quorum is sufficient to conclusively decide a matter, rendering any subsequent public meeting merely *pro forma*, a restriction on meetings of a quorum of a body is narrowly tailored in a way that a restriction on private chance conversations between any two members is not.[144]

In upholding open meetings laws, state courts often simply conclude, without significant discussion, that open meetings laws do not violate free speech rights.[145] One response they give is that, quite the contrary, open meetings laws *promote* free speech, by giving the public an adequate opportunity to participate in public debate.[146] Another approach is to reason that by requiring public notice for discussion of public issues, such laws do not restrict the content of an official's speech, but merely its "location and

---

140. *Id.*

141. *Id.* at 289.

142. *See* Cole v. State, 673 P.2d 345, 350 (Colo. 1983) (applying open meetings law to a meeting of a full legislative caucus); St. Cloud Newspapers, Inc. v. Dist. 742 Cmty. Schs., 332 N.W.2d 1, 7 (Minn. 1983) (series of secret meetings of full membership of the government board)); Sandoval v. Bd. of Regents Univ., 67 P.3d 902 (Nev. 2003) (meeting before a quorum of government body); Smith v. Pa. Employees Benefit Trust Fund, 894 A.2d 874 (Pa. 2006) (meeting before a quorum of government body); Hays Cnty. Water Planning P'ship v. Hays Cnty., 41 S.W.3d 174 (Tex. Ct. App. 2001) (meeting before a quorum of government body).

143. *See Cole*, 673 P.2d at 350; *St. Cloud Newspapers, Inc.*, 332 N.W.2d at 7; *Sandoval*, 67 P.3d at 902; *Smith*, 894 A.2d at 880; *Hays Cnty.*, 41 S.W.3d at 182.

144. *See Cole*, 673 P.2d at 350; *St. Cloud Newspapers, Inc.*, 332 N.W.2d at 7; *Sandoval*, 67 P.3d at 902; *Smith*, 894 A.2d at 880; *Hays Cnty.*, 41 S.W.3d at 182.

145. *See, e.g.*, *Cole*, 673 P.2d at 350 (stating that the statute properly balanced free speech concerns against the public's right of access); *St. Cloud Newspapers, Inc.*, 332 N.W.2d at 7; *Sandoval*, 67 P.3d 902 (dismissing free speech issue in just one sentence); *Smith*, 894 A.2d at 880–81 n.4; *Hays Cnty.*, 41 S.W.3d at 182 (mentioning briefly that the law restricted only the time and place of the speech, and that the officer involved spoke in his official capacity and not as a member of the public). Typical is *St. Cloud Newspapers*, where the state supreme court stated conclusorily that "the legislature is justified in prescribing such openness in order to protect the compelling state interest of prohibiting the taking of actions at secret meetings where the public cannot be fully informed about a decision or . . . detect improper influences." *St. Cloud Newspapers, Inc.*, 332 N.W.2d at 7.

146. *See, e.g.*, *Smith*, 894 A.2d at 880–81 n.4.

timing."[147] Still another defense is that such laws do not restrict an individual's right to speak as a citizen but merely speech in one's capacity as a public official.[148] Or, on a related note, that when one becomes a public official, one forfeits one's right to speak about government affairs in private.[149]

This analysis is incomplete. First, it is not enough to say that because the policy goal of a speech restriction is to foster debate, it survives a free speech challenge.[150] The Supreme Court has stated, "As a matter of constitutional tradition, in the absence of evidence to the contrary, we presume that governmental regulation of the content of speech is more likely to interfere with the free exchange of ideas than to encourage it."[151] For example, campaign finance laws are often defended on the ground that they are designed to level the playing field among donors of varying means, thereby promoting a fair and open debate in elections.[152] Nonetheless, the Supreme Court has subjected laws of this type to exacting scrutiny. Sometimes these laws survive such scrutiny,[153] and sometimes they do not,[154] but courts treat the free speech issues as serious.

Second, open meetings laws do more than merely regulate the "location and timing" of speech.[155] They are not pure "time, place, and manner" regulations but rather laws which impose restrictions based on the content of what is said.[156] This is of course a crucial distinction, inasmuch as "content-neutral" regulations enjoy friendlier treatment by courts.[157] "Content-based" regulations receive "the most exacting scrutiny," known as strict scrutiny, whereas content-neutral regulations receive intermediate scrutiny.[158] Moreover, even if they are properly analyzed as content-neutral restrictions, they are still subject to more than cursory judicial examination.

---

147. *Sandoval*, 67 P.3d at 907; *Hays*, 41 S.W.3d at 182.

148. *See Hays*, 41 S.W.3d at 181–82.

149. State *ex rel.* Murray v. Palmgren, 646 P.2d 1091, 1099 (1982), *appeal dismissed*, 459 U.S. 1081 (1982).

150. Reno v. ACLU, 521 U.S. 844, 885 (1997).

151. *Id.*

152. McConnell v. FEC, 540 U.S. 93, 267 (2003) (citing Buckley v. Valeo, 424 U.S 1, 38 (1976)).

153. *See, e.g.*, FEC v. Colo. Republican Fed. Campaign Comm., 533 U.S. 431, 437 (2001).

154. *E.g.*, Randall v. Sorrell, 548 U.S. 230, 236 (2006) (holding that low, specific ceilings on expenditures violate the First Amendment).

155. *See* ERWIN CHEMERINSKY, CONSTITUTIONAL LAW: PRINCIPLES AND POLICIES 904–08 (2d ed. 2002) (discussing content-based regulation).

156. *Id.*

157. United States v. O'Brien, 391 U.S. 367, 377 (1968).

158. Republican Party of Minn. v. White, 536 U.S. 765, 774–75 (2002); Turner Broad. Sys. v. FCC, 512 U.S. 622, 640 (1994). Under "strict scrutiny," content-based laws are unconstitutional unless the government can show that the law furthers a "compelling governmental interest" and is "narrowly tailored" to further that interest. *White*, 536 U.S. at

B. Content-Based or Content-Neutral?

At this point in the analysis, we should consider whether open meetings laws can truly be considered "content neutral" under applicable Supreme Court First Amendment precedent. The answer is surprisingly unclear.

The "sunshine" laws are not like an ordinance forbidding loud public displays in residential areas after 11 p.m. on weekdays, or a "Post No Bills" sign on the walls of public buildings.[159] Such rules are truly "content-neutral" because the restrictions are the same despite the subject matter of the oral speech or written material involved.[160] The open meetings laws ban only *discussion of official business* outside "sunshined" public meetings.[161]

Further, it is no defense to say that the government is not discriminating in favor of speech on one side of an issue, but rather only forbidding a certain general topic of speech in the proscribed context.[162] A content-based law regulating a certain subject matter is still subject to strict scrutiny even if it is "viewpoint-neutral."[163]

City of Cincinnati v. Discovery Network, Inc. is a good example.[164] Cincinnati banned from city property newsracks containing "commercial handbills" but permitted newsracks containing "newspapers."[165] The law did not discriminate on the basis of the viewpoint expressed by either newspapers or commercial bills.[166] The Supreme Court analyzed the ordinance as content-based, stating that "whether any particular newsrack falls within the ban is determined by the content of the publication resting inside that newsrack. Thus, by any commonsense understanding of the term, the ban is 'content-based.'"[167] This analysis is representative of the Court's approach in these cases.[168] If liability under the law depends on the

---

774–75. By contrast, the "intermediate scrutiny" applied to content-neutral laws requires only an "important governmental interest" to justify the law; the law must only be "substantially related" to furthering that interest. *O'Brien*, 391 U.S. at 377.

159. CHEMERINSKY, *supra* note 155, 904–09.

160. *Id.*

161. SCHWING, *supra* note 1, at 23–24.

162. *See* Hill v. Colorado, 530 U.S. 703, 723–24 (2000); Carey v. Brown, 447 U.S. 455, 462 n.6 (1980); Consol. Edison Co. v. Pub. Serv. Comm'n, 447 U.S. 530, 538 (1980).

163. *Hill*, 530 U.S. at 723; *Carey*, 447 U.S. at 462; *Consol. Edison Co.*, 447 U.S. at 538.

164. 507 U.S. 410 (1993).

165. *Id.* at 414–15.

166. *Id.* at 431.

167. *Id.* at 429.

168. *See, e.g.*, Turner Broad. Sys. v. FCC, 512 U.S. 622, 648 (1994) (treating as content-neutral a federal law requiring cable TV channels to carry local broadcast stations because it included all broadcast stations regardless of the content of the stations' programs); Ward v. Rock Against Racism, 491 U.S. 781, 792 (1989) (treating as content-neutral an ordinance regulating sound levels at public concerts because it applied equally to all types of speech and music).

content of the speech in question, it will very likely be treated as a "content-based" restriction.[169] By this logic, open meetings laws ought to be considered content-based regulations and subjected to "strict scrutiny" analysis.

However, there are reasons for doubting this conclusion. The Supreme Court often states that an important factor in classifying a speech restriction as content-based is whether the government imposes the restriction "because of disapproval of the ideas expressed."[170] Preventing this type of censorship is the core value underlying the Court's special hostility toward content-based regulations.[171] An alternative formulation is that content-neutral regulations are "justified without reference to the content of the speech,"[172] or that with such regulations, there is "no realistic possibility that official suppression of ideas is afoot."[173] A court evaluating a free speech challenge to an open meetings law could very easily conclude that its goal is not "official suppression of ideas," nor is it motivated because of disapproval of the ideas expressed in the covered speech.[174] These conclusions would argue for treating the law as content-neutral.[175] Indeed, the district court in the recent *Asgeirsson* case so found.[176]

A closer question is whether open meetings laws impose a restriction "because of disapproval of the ideas expressed."[177] While the governing body passing an open meetings law may not disapprove of the specific content of any particular statement made among public officials outside of a "sunshined" meeting, it undoubtedly "disapproves" of the expression of those ideas (and only those ideas) in such a context in the first place.

A leading case on this point is *Renton v. Playtime Theatres, Inc.*[178] In *Renton*, the Supreme Court upheld a local zoning ordinance, which prevented an adult movie theatre (i.e., one showing sexually explicit content) from locating within 1,000 feet of schools, churches, and certain residential areas.[179] Because the ordinance plainly affected only sexually explicit movies, it was undoubtedly "content-based" in a literal sense.[180] But the Court held that it did "not fit neatly into either the 'content-based'

169. *Turner Broad.*, 512 U.S. at 642.

170. R.A.V. v. City of St. Paul, 505 U.S. 377, 382 (1992).

171. *Id.*

172. *Id.* at 389 (citing Renton v. Playtime Theatres, Inc., 475 U.S. 41, 48 (1986)).

173. *Id.* at 390. On the other hand, the Court has also cautioned that a content-discriminatory purpose is *sufficient,* but not *necessary*, to show that a law is content-based. *Turner Broad.*, 512 U.S. at 642.

174. *R.A.V.*, 505 U.S. at 389–90.

175. *Id.*

176. Asgeirsson v. Abbott, No. P-09-CV-59, 2011 WL 1157624, at 14–15 (W.D. Tex. Mar. 25, 2011).

177. *Id.* at 382.

178. 475 U.S. 41 (1986).

179. *Id.* at 54–55.

180. *Id.* at 57–58.

or 'content-neutral' category"[181] because the law was "aimed not at the *content* of the films . . . but rather the secondary effects of such theaters on the surrounding community."[182] The Court relied on a district court finding that the "predominant motive" of the local body passing the law was to prevent the negative effects these theaters had on the surrounding neighborhoods with respect to crime, property values, and the retail trade.[183] Thus, the law should be treated as content-neutral.[184] Citing *Renton,* the Supreme Court has used this "secondary effects" analysis to treat as content-neutral other laws that would be considered content-based under a more literal approach.[185] By analogy, then, open meetings laws may be analyzed as "content-neutral" in this sense. Again, the district court in *Asgeirsson* so held.[186]

The mere articulation of "secondary effects" by a defendant government entity, however, is not enough to switch all literally content-based laws to the more lenient content-neutral treatment. In *City of Cincinnati,* for example, the city tried to rely on *Renton* by arguing that its newsrack ordinance was motivated by the content-neutral concerns of safety and aesthetics related to overcrowding of public spaces.[187] The Court rejected this argument, noting that these supposed "secondary effects" were not more related to "commercial handbills" than newspapers, and thus did not justify an ordinance banning all commercial handbills but allowing newspapers.[188]

Similarly, in *Boos v. Barry,*[189] the Supreme Court struck down a law banning protests critical of a foreign government within 500 feet of the government's embassy.[190] The Court rejected a *Renton* analogy for a somewhat different reason, emphasizing that the "secondary effects" cited by the government—shielding foreign diplomats from speech offending their dignity—was *related to the content of the speech.*[191] This is in accord with Supreme Court precedent generally, which requires that the

181. *Id.* at 47 (emphasis added).

182. *Id.* (emphasis in original).

183. *Id.* at 48.

184. *Id.*

185. *See, e.g.*, Hill v. Colorado, 530 U.S. 703, 720 (2000) (concerning a state law banning approaching within eight feet of a person who is within 100 feet of a health facility for purposes of "protest, education, or counseling"); City of Erie v. PAP's AM, 529 U.S. 277, 294–95 (2000) (involving a city ordinance barring public nudity).

186. Asgeirsson v. Abbott, No. P-09-CV-59, 2011 WL 1157624, at 14–15 (W.D. Tex. Mar. 25, 2011).

187. 507 U.S. at 430.

188. *Id.*

189. 485 U.S. 312 (1987).

190. *Id.* at 337–38.

191. *Id.* at 320–21.

governmental interests articulated to justify an assertedly content-neutral speech restriction be "unrelated to the suppression of free expression."[192]

The Supreme Court cases in this area are not entirely clear on how to tell the difference between a content-based and content-neutral standard.[193] One useful way to synthesize the different cases in this area is to say there is a presumption that laws explicitly referencing a particular subject matter or content of speech will be treated as content-based. This presumption may be overcome if the defendant can show that the law is aimed at a "secondary effect" unrelated to the content of the speech.[194] However, the presumption may only be rebutted if the court is convinced that the secondary effects are related to the banned category of speech and not equally related to the permitted category of speech.[195]

Under this analysis, open meetings laws are presumptively content-based and thus presumptively subject to strict scrutiny. A government defending such a law against a free speech challenge would have a reasonable argument in rebuttal that the law is aimed not at the content of the speech but at the "secondary effect" of excluding the public from debate leading to decisions by their government representatives. This secondary effect is clearly present with the banned category of speech—discussion of action to be taken by the government body—and not present with the unbanned categories of speech: all other speech.

The closer question is whether this secondary effect is truly unrelated to the content of the speech. One can characterize the government's purpose here as keeping the public involved in the debate (content-neutral) but doing so by stifling any discussion by covered officials of relevant public policy issues (content-based).[196] Are open meetings laws more like the content-neutral zoning restriction on adult theaters in *Renton*, and thus, to be treated as effectively content-neutral? Or are they more like the content-based restriction on opposition protests near foreign embassies in *Boos*?

Two useful analogous Supreme Court cases point in opposite directions on this question.[197] In *Colorado v. Hill*,[198] a state law barred anyone from approaching within eight feet of a person who was within 100 feet of a health care facility. The law specifically barred such approaches only when done with the purposes of "oral protest, education, or counseling," which arguably suggests a content-based law.[199] Nonetheless, the Court analyzed

192. Texas v. Johnson, 491 U.S. 397, 410–13 (1989) (citing United States v. O'Brien, 391 U.S. 367, 377 (1968)).

193. *See* CHEMERINSKY, *supra* note 15555, at 908–09 (citing United States Supreme Court cases on the issue).

194. *See id.*

195. *Id.*

196. *See id.* at 904–09 (discussing content-based and content-neutral regulations).

197. Hill v. Colorado, 530 U.S. 703, 703 (2000); Carey v. Brown, 447 U.S. 455, 455 (1980).

198. 530 U.S. 703 (2000).

199. *Id.* at 720.

the law as content-neutral.[200] The Court took pains to distinguish its decision in *Carey v. Brown*,[201] where it struck down as content-based an Illinois law banning picketing that contained an exemption for picketing a place of employment involved in a labor dispute. In contrast, the Court in *Hill* reasoned that the Colorado law "simply establishes a minor place restriction on an extremely broad category of communications with unwilling listeners."[202] Although perhaps inspired by abortion protests, the Colorado law applied equally to protests or other communications regarding animal rights, the environment, or any other subject.[203]

Further, the Colorado law was not objectionably under-inclusive in terms of the types of speech it covered. As the Court explained, a speech restriction only "lends itself to invidious use if there is a significant number of communications, raising the same problem that the statute was enacted to solve, that fall outside the statue's scope, while others fall inside."[204] The even-handedness of the constitutionally valid Colorado law stands in contrast with the fatal under-inclusiveness of the Illinois picketing ban's exemption for labor disputes.

Applied to open meetings laws, the analysis in *Hill* argues for a content-neutral label. Although such laws do explicitly restrict a particular topic of speech, they arguably involve "a minor place restriction on an extremely broad category of communication," designed in this case not to protect "unwilling listeners" but to prevent exclusion of willing listeners. It is arguably either a "minor place restriction" or "minor time restriction," depending on how one views the notion of "outside of a properly noticed public meeting."

Further, the open meeting restriction is arguably not under-inclusive. The category of speech covered—substantive discussion of action by a governmental body by members of that body—leaves out no speech that implicates the asserted governmental interest of including the public in governmental decisions.[205]

A counterargument is that open meetings laws generally are under-inclusive in that they do not cover deliberations by local mayors, elected sheriffs, elected trustees, and other local elected officials, whose decisions often matter far more to average citizens.[206] If the legitimate state interest justifying open meetings laws is to ensure that the public has meaningful access to and input in decisions made by local elected officials, then such laws really are under-inclusive.

---

200. *Id.* at 721.

201. 447 U.S. 455 (1980).

202. *Hill*, 530 U.S. at 723.

203. *Id.* at 723–24.

204. *Id.* This latter point sounds much like narrow tailoring, the second prong of strict scrutiny analysis.

205. SCHWING, *supra* note 1, at 23–24.

206. *See infra* Section V.

Of course, these situations involve single-office elected officials conferring with each other, as opposed to fellow elected members of a joint, collegial elected body deliberating in private. But how persuasive is this distinction? In states where judges are elected, multi-judge judicial panels may have elected judges deliberating in private as they decide cases, yet they are *expected* to deliberate in private. Indeed, if a legislature were to attempt to require multi-judge panels to deliberate publicly, it would be unsurprising to see fellow judges quickly striking down such a law.[207] Defenders of judicial prerogatives would say that such private deliberation is essential for candid discussion, proper outcomes, and the integrity of the decision-making process.

Why is the same not true for legislators? The answer cannot be that judges decide individual cases affecting the legal interests of individual citizens, some of whom may have privacy interests, because many state open meeting laws require legislators to deliberate publicly when they adjudicate personnel grievances, student appeals, and the like. More convincing is the response that judges, unlike legislators, must decide cases based on the law rather than public opinion. But even this is not a complete answer, for where judges are elected, they are elected, at least in part, based on an expectation that their decision making will in some sense reflect public values.

Further, many state legislatures exempted themselves in passing open meetings laws. Given that the "public access and input" rationale applies equally to state legislators as local legislators,[208] all such laws are substantially under-inclusive.

However, all these types of under-inclusion are arguably unrelated to the content of the speech involved and perhaps are distinct from the labor dispute exemption relied upon by the Supreme Court in *Carey v. Brown*. Unless open meetings laws' failure to include deliberations by state legislators where not covered or deliberations by non-legislative elected officials renders them "under-inclusive" in the *Carey* sense, *Colorado v. Hill* suggests that "sunshine" laws should be analyzed as content-neutral regulations.

However, *Burson v. Freeman*[209] seems to counter this suggestion. It is similar to *Hill* but has one key difference—a difference present with open meetings laws—which renders it content-based in the eyes of the Court. *Burson* involved a free speech challenge to Tennessee's law banning

---

207. In some cases, a court might strike down such a law on separation of powers grounds, ruling that the legislature was inappropriately intruding on the independence of the judicial branch. However, not all states' separation of powers doctrines are identical to the federal government's or to each other. If it were somehow necessary or desirable to resolve such a case by resorting to First Amendment principles, it is not hard to imagine fellow judges doing so to protect judicial prerogatives.

208. *See infra* Section IV.

209. 504 U.S. 191 (1992).

solicitation of votes and display of campaign materials within one hundred feet of a polling place. It was similar to the ordinance at issue in *Colorado v. Hill*, except that it did not ban *any* approach of a person within one hundred feet of a polling place, but only those involving solicitation of votes.

Because the applicability of the statute depended on the subject matter of what was to be discussed, as well as the physical location, the Supreme Court flatly rejected the State's argument that it was a content-neutral "time, place, or manner" restriction.[210] The Court explained that this must be so because "[w]hether individuals may exercise their free speech rights near polling places depends entirely on whether their speech is related to a political campaign."[211] The Court then held that it must apply strict scrutiny because it was a content-based speech restriction.[212] The Court eventually upheld the restriction as narrowly tailored to further the compelling governmental interests of protecting against voter intimidation and election fraud.[213]

Burson is strikingly similar to the case of open meetings laws. In both cases, to protect the interests of voters, the state imposed a restriction on speech that depended both on the time and place of the speech: within one hundred feet of a polling place or outside of a publicly noticed public meeting. The application of the speech restriction depended additionally on the topic of the speech itself: political campaign speech or substantive discussion of local government business. As the Court explained in *Burson*, the statute "implicates three central concerns in our First Amendment jurisprudence: regulation of political speech, regulation of speech in a public forum, and regulation based on the content of the speech."[214]

Another analogous situation is *Republican Party of Minnesota v. White*,[215] where the Court struck down a state supreme court's judicial canon preventing judicial election candidates from announcing their views on disputed legal or political issues. The Court concluded that the rule was indeed a content-based restriction subject to strict scrutiny.[216] The content-based nature of the rule was apparently not disputed, and the Court seemed to think it obvious that the rule should be so characterized. It did note that the rule was under-inclusive because it was limited to the time period of a judicial election campaign but not to the periods before or after, unless a specific case regarding the legal or political issue in question was pending.[217]

---

210. *Id.* at 197–98.
211. *Id.* at 197.
212. *Id.*
213. *Id.* at 207–10.
214. *Id.* at 196.
215. 536 U.S. 765 (2002).
216. *Id.* at 774–75.
217. *Id.* at 787–88.

Like the restriction in *White*, the open meetings law was a restriction placed on a public official which prevented the official from discussing a large category of public issues related to his office during a specified time period. In the case of sunshine laws, the time period is "any time other than at a properly noticed public meeting;" in the *White* case, it was "during a judicial election campaign." Unlike the restriction in *White*, the open meetings law was not under-inclusive. Nevertheless, there are sufficient similarities with *White* to suggest that an open meetings law might be properly analyzed as content-based and thus subject to strict scrutiny.

The only federal appellate court to have considered whether open meetings laws are content-neutral characterized them unqualifiedly as content-based regulations subject to strict scrutiny. In *Rangra v. Brown*,[218] the Fifth Circuit considered an appeal from an elected official bringing a free speech challenge to Texas' open meetings law. Relying on *White* and *Burson*, the court concluded that the law was indeed a content-based speech restriction subject to strict scrutiny.[219]

Indeed, the Fifth Circuit cited language in Supreme Court cases establishing that regulation of *political* speech would normally trigger strict scrutiny.[220] This notion is in accord with First Amendment doctrine, which states generally that protection of political speech and discussion of public issues is a central value of the First Amendment, one affording such speech heightened protection.[221] Thus, the only Circuit-level case to have explicitly discussed the proper standard of review for a free speech challenge to an open meetings law has held that the strict scrutiny standard of content-based regulations applies.

1. Applying Strict Scrutiny

If open meetings laws are indeed content-based, there is a very good chance that some of the more broad-reaching provisions of such laws may be successfully challenged. Content-based speech restrictions will fail the

---

218. 566 F.3d 515 (5th Cir. 2009).

219. *Id.* at 518, 520–25. Because this case was later dismissed as moot, the *Rangra* decision lacks formal precedential value. Nonetheless, it provides significant guidance as the only federal circuit court case to have considered the question.

As this article goes to press, plaintiffs are appealing (see note 24) the recent district court decision which acknowledged this Fifth Circuit holding, noted that the Fifth Circuit hold no longer has precedential value, and held that the Texas open meetings law was content-neutral. *See* Argeirsson v. Abbott, No. P-09-CV-59, 2011 WL 1157624 (W.D. Tex. Mar. 25, 2011). The court reasoned that, *inter alia*, the law was unrelated to the suppression of speech and targeted "secondary effects." *Id.* at 14-15.

220. *See id.*

221. *See* Connick v. Myers, 461 U.S. 138, 145 (1983) ("The First Amendment 'was fashioned to assure unfettered interchange of ideas for the bringing about of political and social changes desired by the people.'") (quoting Roth v. United States, 354 U.S. 476, 484 (1957)); N.Y. Times v. Sullivan, 376 U.S. 254, 269 (1964).

"strict scrutiny" test unless the government can show that they are "narrowly tailored" to serve "a compelling [g]overnment interest."[222] To be narrowly tailored, the law must be the least restrictive means available to serve the compelling governmental interest.[223] If another, less restrictive provision would serve the governmental interest equally, the legislature must use such a provision.[224] Indeed, if a plaintiff proffers any alternative provision, then the burden is on the government to prove that the proposed alternatives will not be as effective as the challenged statute.[225]

For example, it seems a stretch to say that broad laws reaching any substantive conversation between any two legislators are narrowly tailored. Indeed, most open meetings laws do not reach this broadly. Instead, they bar a *quorum* of a legislative body from secretly discussing a pending matter.[226] In such cases, there is a danger that the public will be shut out of meaningful participation in the decision-making process. A quorum could decide on a course of action in advance, then meet in a *pro forma* public meeting where the preordained conclusion is "rubber-stamped." In any typical-sized legislative body, a conversation between two, or even three, legislators poses no such realistic danger.[227] Other states strike a middle ground of barring a majority of a quorum from discussing matters privately.[228]

There is no evidence to suggest that democracy is significantly impaired in these more permissive states, which constitute the overwhelming majority. Thus, while a "quorum rule" seems constitutionally defensible, and a "half a quorum rule" provides a closer case, it is much harder to characterize as "narrowly tailored" a broad, Tennessee-style rule preventing any two legislators from ever having a substantive discussion about government decisions outside a properly noticed public meeting.

Similarly, one could plausibly argue that narrow tailoring would require exceptions for discussion of sensitive matters for which legislators would

---

222. United States v. Playboy Entm't Grp., Inc., 529 U.S. 803, 813 (2000).

223. *Id.*

224. *Id.*

225. Ashcroft v. ACLU, 542 U.S. 656, 665 (2004).

226. SCHWING, *supra* note 1, at 265.

227. One objection to this argument is that allowing two legislators to confer outside a publicly noticed meeting can "open the floodgates." Legislator A could confer separately with Legislator B and C, while Legislator D confers separately with Legislator E and F, thus allowing a final conference between Legislators A and D to accomplish the equivalent of a quorum meeting. However, most states that use a "quorum rule" or "half a quorum rule" expressly ban the use of such serial communications to accomplish indirectly what cannot be accomplished directly. *See, e.g.*, Sutter Bay Assocs. v. Cnty. of Sutter, 68 Cal. Rptr. 2d 492 (Cal. Ct. App. 1997); Moberg v. Indep. Sch. Dist., 336 N.W.2d 510 (Minn. 1983).

228. *E.g.*, 5 ILL. COMP. STAT. 120/1.02 (West 2010); KAN. STAT. ANN. § 75-4317(a) (West 2008); *see* SCHWING, *supra* note 1, at 271 (showing a further examination of compromises defining meetings through partial quorum counts).

have a legitimate desire to discuss in private. Examples might include personnel matters, matters that involve individual citizens' privacy, or consultations with government counsel over pending legal matters. They might also include competitive financial negotiations between the government entity and an outside party, whether it be collective bargaining with a local union, negotiations with a potential vendor, or discussions of the proper price for which a local government might sell government-owned land or acquire new land from private owners. Various states have exemptions to their open meetings laws covering precisely such areas, but there are many states which recognize only a few or none of these exceptions.[229]

Aside from requiring the least restrictive burden on speech, the narrow tailoring requirement also guards against "under-inclusive" speech restrictions. In *R.A.V. v. City of St. Paul*,[230] the Supreme Court explained that even where the State regulates a category of speech previously ruled to be unprotected by the First Amendment, such as obscenity, that regulation may run afoul of the First Amendment if it is "under-inclusive"—that is, it regulates only some of the unprotected speech but not all of it.[231] In *R.A.V.*, the Court struck down a "hate crimes" ordinance that made it an offense to display any symbol while knowing or having reason to know that it "arouses anger, alarm or resentment in others on the basis of race, color, creed, religion or gender."[232] The ordinance was unconstitutional on the grounds that its protection from fear or alarm was limited to narrow classes of speech.[233] Significantly, while the Court discussed the ordinance's potential viewpoint discrimination by noting that "[o]ne could hold up a sign saying, for example, that all 'anti-Catholic bigots' are misbegotten; but not that all 'papists' are, for that would insult and provoke violence 'on the basis of religion,'"[234] it went further, suggesting that even non-viewpoint-discriminatory under-inclusiveness might also invalidate such a law.[235] Thus, while a State could ban all obscenity, it could not ban just obscenity offensive to African-Americans.

This notion of fatal under-inclusiveness is not limited to the regulation of unprotected speech. For example, in *Carey v. Brown*, the Court cited the under-inclusiveness of a law which barred picketing but exempted labor

229. *Compare* ALA. CODE § 36-25A-7 (2010) (providing numerous enumerated exceptions), *and* MISS. CODE ANN. § 25-41-1 (West 2010), *with* ALASKA STAT. ANN. § 44.62.310 (West 2009) (providing exceptions only in cases pertaining to personal character or information made secret by statute), *and* ARK. CODE ANN. § 10-3-305 (West 2010).

230. 505 U.S. 377 (1992).

231. *See id.* at 377.

232. *Id.* at 380, 391.

233. *Id.* at 391.

234. *Id.* at 391–92.

235. *Id.* at 387.

disputes, characterizing the law as content-based and thus subject to the most exacting scrutiny.[236]

This is another significant constitutional vulnerability of the broadest of the open meetings laws. Those that exempt state legislatures are exceedingly under-inclusive. Additionally, open meetings laws do not reach consultations involving single-office elected officials, such as mayors, sheriffs, and trustees, or between one or more of them and a local legislator. Such under-inclusivity is substantial.

2. Applying Intermediate Scrutiny

Even if open meetings laws are properly characterized as content-neutral, a significant amount of judicial examination is still required. A court would still apply "intermediate scrutiny." Under this standard, the government would be required to show (1) that the law "furthers an important or substantial governmental interest;" (2) that the interest is "unrelated to the suppression of free expression;" (3) that "ample alternative channels" for communication of the information exist; and (4) that the incidental restriction on alleged First Amendment freedoms is no greater than is essential to the furtherance of that interest."[237] To satisfy this last criterion, the regulation need not be the least speech-restrictive means of advancing the government's interests. Here, "narrow tailoring" is satisfied if the means chosen do not "burden substantially more speech than is necessary to further the government's legitimate interests."[238]

Regardless of whether the state interest of preserving public access to the decision-making process is "compelling," it is at least "substantial," and, as noted above, it is unrelated to suppression of free expression. This standard's effect on open meetings laws thus turns on the application of prongs (3) and (4).

For a covered government official who wishes to consult with colleagues on a governmental matter, there are very few "alternative channels" available. The government official can either consult with those colleagues in a properly noticed public meeting, or, in most states, make a public statement to the media. In some states, the government official could not even circulate a "Dear Colleague" letter outlining the official's position outside a publicly noticed meeting even if the official were to copy the local media on it. A serious question arises as to whether this limited menu of alternative channels is "ample."

The hypotheticals that began this Article illustrate the point. Consider the County Commissioner who wishes to email colleagues a detailed memo

---

236. 447 U.S. 455 (1980).

237. *See* Ward v. Rock Against Racism, 491 U.S. 781, 791 (1989); Heffron v. Int'l Soc'y for Krishna Consciousness, 452 U.S. 640, 649–55 (1981).

238. Turner Broad. Sys. v. FCC, 512 U.S. 622, 642 (1994) (citing *Ward*, 491 U.S. at 799).

analyzing a draft ordinance suggesting draft amendatory language for consideration prior to the next County Commission meeting. Local media will not generally oblige the Commissioner by printing such memos for all the world to see, and scheduling a publicly noticed "pre-meeting meeting" will in most cases be impractical for reasons of time and colleagues' availability. Ditto for the Democratic and Republican legislators who seek to meet out of committee to craft compromise language to settle a sizzling partisan dispute.

Or consider the alderperson who attends a public forum on an urgent local issue and who wishes to engage in the debate on that issue. If the meeting is at an organization not open to the whole public—e.g., a local party caucus, or a dues-based membership organization like the Jaycees—or even if the meeting is open to all but was simply not properly "sunshined" in accordance with the local open meetings law, the alderperson should hope that no colleague from the aldermanic council is present in the audience. If a colleague is present, then both are under an effective gag order. In these situations, the "alternative channels" available under most open meetings laws simply do not afford the officials a practical manner to convey their views or seek the views of colleagues. These channels hardly sound "ample."

Regarding the fourth prong, there is likewise a significant issue as to whether the typical open meetings law "burdens substantially more speech than is necessary" to further the government's legitimate interest. While the government has a legitimate interest in assuring public access to legislative decision making as a general matter, it is by no means clear that that interest extends to ensuring that legislators do not have the ability to confer collectively with counsel in private regarding pending legal matters, or to discuss in private sensitive personnel matters or threats to individual privacy. Nor is it clear that this interest extends to preventing legislators from conferring with each other about what negotiating position they should take with (1) an outside vendor seeking a government contract; (2) a union conducting collective bargaining; (3) a landowner hoping to sell land to the government; or (4) a potential purchaser negotiating the purchase of government-owned land. Finally, it is questionable how significant a public interest there is in barring legislative leaders from either party from ever meeting privately to broker a compromise on a difficult public policy question. If anything, the government interest seems to point in the opposite direction for each of these examples. If even some of these cases are examples of speech banned by open meetings laws without a legitimate government interest, these bans would limit "substantially" more speech than necessary.

C. Legislators as Public Employees

State courts have also dismissed free speech challenges to open meetings laws because such laws do not regulate individuals' speech as

private citizens: instead, the laws cover their speech as officials.[239] For example, in *Hays County Water Planning Partnership v. Hays County*, for example, the Texas Court of Appeals noted that the types of statements covered by the open meetings law would be made by plaintiff Commissioner as a Commissioner and not during the "public comment" portion of the meeting, when each citizen is given three minutes to speak.[240] Support for this approach arguably can be derived from a series of Supreme Court cases establishing lower free speech protections for government employees.[241] However, as the Fifth Circuit has recognized, these cases are inapposite, and government officials' free speech rights are not subordinate to those of others in the open meetings law context.

The most recent public employee case is *Garcetti v. Ceballos*.[242] In *Garcetti*, the Supreme Court held that the First Amendment does not protect a government employee from discipline based on speech made pursuant to the employee's official duties. The case involved a deputy district attorney who wrote an internal office memorandum criticizing law enforcement's handling of a case and recommending dismissal.[243] The deputy district attorney was later subject to adverse employment actions, claimed retaliation, and brought a First Amendment claim.[244] The Court dismissed the claim, holding that the speech involved was not subject to First Amendment protection.[245] The Court explained that when public employees speak as part of their official duties, they "are not speaking as citizens for First Amendment purposes . . . ."[246]

The Court cited its earlier decision in *Connick v. Myers*,[247] upholding a decision to discipline a public employee for writing and distributing an internal office questionnaire devoted mostly to internal issues of office morale and reassignment policies.[248] In *Connick*, the Court held that public employees were entitled to protection for speech "made as a citizen on matters of public concern" but not for speech made "as an employee on matters only of personal interest."[249] In *Garcetti*, the Court clarified that even if the public employee's speech concerned a "matter of public concern," it would not qualify as being made "as a citizen" if the speech

239. *See* Hays Cnty. Water Planning P'ship v. Hays Cnty., 41 S.W.3d 174, 182 (Tex. Ct. App. 2001).

240. *Id.*

241. *See id.* (citing Connick v. Myers, 461 U.S. 138, 147 (1983); Pickering v. Bd. of Educ., 391 U.S. 563 (1968)).

242. 547 U.S. 410 (2006).

243. *Id.* at 413–17.

244. *Id.* at 415.

245. *Id.* at 417.

246. *Id.* at 421.

247. 461 U.S. 138 (1983).

248. *See id.*

249. *Id.* at 147.

were made as part of the discharge of the employee's official duties.[250] In sum, the "government employee" cases hold that a public employee's speech is protected under the First Amendment only when it (1) involves a matter of public concern, and (2) was made in the individual's capacity as a citizen, not as part of the employee's duties.

Drawing upon *Garcetti*, one could argue that those subject to open meetings laws cannot raise a free speech challenge because when covered by the laws, they are not speaking as a "citizen" but as an official as part of their official duties. This was indeed the track taken by the district court in *Rangra*. The trial court had rejected the free speech challenge, holding that after *Garcetti*, the First Amendment affords no protection to speech by elected officials made pursuant to their official duties.[251]

However, this analysis is also suspect, as the Fifth Circuit made clear in its overruling of the *Rangra* trial court.[252] The key lies in the reason behind the lesser protections afforded public employees in the first place. As the Fifth Circuit explained, job-related speech by public employees is less protected[253] than other speech because employee speech rights must be balanced with "the government's need to supervise and discipline subordinates for efficient operations."[254]

In these public employee cases, the Supreme Court has repeatedly made clear that government has more power to restrict speech when it acts as an employer supervising an employee as opposed to a sovereign writing rules for persons generally. In *Pickering v. Board of Education*,[255] the Court stated that "[t]he State has interests as an employer in regulating the speech of its employees that differ significantly from those it possesses in connection with regulation of the speech of the citizenry in general."[256] In *Garcetti* itself, the Court stated, "The government as employer indeed has far broader powers than does government as sovereign."[257]

This is the case because, in order to function effectively, government officials must be able to supervise and discipline their employees and make judgments about their work performance based on, among other things, statements they make at work. As the majority in *Garcetti* put it, "Supervisors must ensure that their employees' official communications are

250. *Garcetti*, 547 U.S. at 421–23.

251. Rangra v. Brown, 566 F.3d 515, 517–18 (5th Cir. 2009).

252. *See id.*

253. I say "less protected" because the Court has made clear that even where the public employee speech is not made "as a citizen" or on "a matter of public concern," it is not *completely* without First Amendment protection. For example, if such an employee were to be sued for defamation, the same First Amendment protections afforded all defamation defendants would still apply. *See* Connick v. Myers, 461 U.S. 138, 147 (1983).

254. *Rangra*, 566 F.3d at 522.

255. 391 U.S. 563 (1968).

256. *Id.* at 568.

257. Garcetti v. Ceballos, 547 U.S. 410, 418 (2006) (quoting Waters v. Churchill, 511 U.S. 661, 671 (1994)).

accurate, demonstrate sound judgment, and promote the employer's mission."[258] The Court was concerned that if the rule were otherwise, every employer-employee dispute could potentially wind up in federal court, and it did not want to "constitutionalize the employee grievance."[259]

Once we consider the underlying reasons for the *Garcetti/Connick* rule limiting government employees to First Amendment protection only for speech made as a citizen, the analogy to sunshine laws weakens substantially. After all, elected officials are not subject to the type of employer discipline relevant to *Garcetti* and its predecessors. In *Rangra*, the Fifth Circuit held:

> While *Garcetti* added a new qualification of public employees' freedom of expression recognized by the Court's long line of cases concerning public employee speech rights, it did nothing to diminish the First Amendment protection of speech restricted by the government acting as a sovereign rather than as an employer and *did nothing to impact the speech rights of elected officials whose speech rights are not subject to employer supervision or discipline.*[260]

A district court applying *Garcetti* has made a similar distinction between government officials and government employees, noting that the "bureaucratic concerns" regarding employee discipline and supervision simply did not apply to local elected or appointed officials.[261]

Indeed, as the Fifth Circuit noted in *Rangra*, case law is clear that First Amendment protection of elected officials' speech is "robust and no less strenuous than that afforded to the speech of citizens in general."[262] *White*, the Supreme Court case discussed above, is a recent example.[263] Invalidating the restrictions on judicial candidates' comments, the *White* Court reaffirmed that "[t]he role that elected officials play in our society makes it *all the more imperative* that they be allowed freely to express themselves on matters of current public importance."[264] As the *Rangra*

---

258. *Id.* at 422 ("The fact that his duties sometimes required [plaintiff] to speak or write does not mean his supervisors were prohibited from evaluating his performance."); *see also Waters*, 511 U.S. at 675 ("When someone who is paid a salary so that she will contribute to an agency's effective operation begins to do or say things that detract from the agency's effective operation, the government employer must have some power to restrain her.").

259. *Garcetti*, 547 U.S. at 420 (quoting Connick v. Meyers, 461 U.S. 138, 154 (1983)).

260. Rangra v. Brown, 566 F.3d 515, 523–24 n.23 (5th Cir. 2009) (emphasis added).

261. *See* Conservation Comm'n of Westport v. Beaulieu, No. 01-11087-RGS, 2008 WL 4372761, at *4 (D. Mass. Sept. 18, 2008) ("Although the Selectmen are the appointing body and have the power to remove the Commissioners for cause, they do not have supervisory authority or managerial control over the Commissioners' day-to-day activities.").

262. *Rangra*, 566 F.3d at 524.

263. Republican Party of Minn. v. White, 536 U.S. 765, 774–75 (2002).

264. *White*, 536 U.S. at 781–82 (emphasis added) (quoting Wood v. Georgia, 370 U.S. 375, 395 (1962)); *see also* Bond v. Floyd, 385 U.S. 116, 133–35 (1966) (reinstating a state representative excluded from the legislature because of his statements criticizing the

court noted, there is no shortage of cases upholding the free speech rights of elected officials and candidates.[265]

For example, in *Eu v. San Francisco County Democratic Central Committee*,[266] a state law purported to prevent local political party officials from endorsing candidates in primary elections.[267] The law also barred such candidates from claiming their party's endorsement in a primary election.[268] Certainly, there was a "good government" state interest there: it should be up to the primary voters to decide which candidate deserves the party's nomination, and the party endorsement may be seen as an unfair advantage for party "insiders" in such a contest. Nonetheless, the Supreme Court invalidated the law, holding both that party officials have a First Amendment right to endorse candidates in primary elections and that such candidates have a First Amendment right to claim party endorsement.[269] Also, in *Brown v. Hartlage*,[270] the Supreme Court held that a candidate has a right to promise to reduce his salary, despite laws banning promises of "any thing of value" in consideration of votes.[271]

Thus, in the *Rangra* case, the Fifth Circuit emphasized that the open meetings law at issue was a content-based restriction on political speech and invalid unless it met strict scrutiny.[272] Reversing the district court decision ruling that the speech in question was outside First Amendment protection, the court remanded the case for application of the strict scrutiny standard.[273] Application of that standard to open meetings laws generally raises serious constitutional doubts about such laws, at least in their most broad form. Even under the more lenient intermediate standard for content-neutral speech restrictions, the broadest of these laws raise significant constitutional issues.

Vietnam War and the draft) (cited in *Rangra*, 566 F.3d at 524). Indeed, the importance of the ability of legislators to speak freely is also reflected in the doctrine of legislative immunity. *See* U.S. CONST., art. I, § 6, cl. 1 (Speech and Debate Clause); Gravel v. United States, 408 U.S. 606, 623–25 (1972); Tenney v. Brandhove, 341 U.S. 367, 379 (1951).

265. *Rangra*, 566 F.3d at 524–25 n.24 (citing Eu v. San Francisco Cnty. Democratic Cent. Comm., 489 U.S. 214, 222 (1989) (upholding the right of party officials to endorse candidates in primary elections, and candidates to claim party endorsement)); *see also* Brown v. Hartlage, 456 U.S. 45, 53–54, 58 (1982) (upholding the right of candidates to promise to reduce their salary, despite laws banning promises of any thing of value in consideration of votes).

266. 489 U.S. 214 (1989).

267. *Id.*

268. *Id.*

269. *Id.* at 222–29.

270. 456 U.S. 45 (1982).

271. *Id.* at 53–54.

272. Rangra v. Brown, 566 F.3d 515, 526–27 (5th Cir. 2009).

273. *Id.*

IV. EQUAL PROTECTION ISSUES

A. Generally

Another potential ground for challenging open meeting laws is equal protection. The Equal Protection Clause of the Fourteenth Amendment provides that "no State shall deny to any person within its jurisdiction the equal protection of the laws."[274] This is "essentially a direction that all persons similarly situated should be treated alike."[275]

The Supreme Court has developed a multi-tiered approach to Equal Protection doctrine. The general rule is that laws creating classifications—i.e., differences in treatment—among different categories of persons will be upheld against an Equal Protection challenge as long as they are "rationally related to a legitimate state interest."[276] This "rational basis" standard of review is a lenient one, requiring validation of challenged laws unless the relationship of the classification to the asserted state interest "is so attenuated as to render the distinction arbitrary or irrational."[277] Generally, social and economic regulation is subject to mere rational basis review,[278] as are classifications based on such categories as class,[279] age,[280] and disability.[281]

In contrast, classifications that burden suspect classes are subject to a heightened form of review.[282] Classifications based on race, alienage, and nationality are subject to "strict scrutiny"—the most exacting form of constitutional review.[283] Such laws will be upheld only if they are "narrowly tailored" to further a "compelling state interest."[284] The governmental interest served must be one of the most fundamental interests served by government, and the means used to serve that end must discriminate against the affected group no more than necessary to achieve the end.[285]

---

274. U.S. CONST. amend. XIV, § 1.

275. City of Cleburne v. Cleburne Living Ctr., 473 U.S. 432, 439 (1985) (citing Plyler v. Doe, 457 U.S. 202, 216 (1982)); *see also* Village of Willowbrook v. Oleck, 528 U.S. 562, 564–65 (2000).

276. *City of Cleburne*, 473 U.S. at 440.

277. *Id.* at 446.

278. *Id.* at 440.

279. San Antonio Indep. Sch. Dist. v. Rodriguez, 411 U.S. 1, 16, 62 (1973).

280. Kimel v. Fla. Bd. of Regents, 528 U.S. 62, 83–84 (2000).

281. Ala. Bd. of Trs. v. Garrett, 531 U.S. 356, 366–67 (2001); *City of Cleburne*, 473 U.S. at 445–46.

282. *Id.* at 440.

283. *Id.*

284. *Id.*

285. *See, e.g.*, Dunn v. Blumstein, 405 U.S. 330, 343 (1972).

Meanwhile, classifications based on gender[286] and illegitimacy[287] are subject to "intermediate scrutiny," and will be upheld as long as the differences in treatment involved are "substantially related to an important governmental objective."[288] This standard requires more than some non-arbitrary, reasonable relationship between the asserted legitimate government interest and the difference in treatment between groups. The involved government interest needs to be more than merely legitimate: it must be "important." Further, the classification involved, while not necessarily the most narrow possible to achieve the important end, must not involve significant under-inclusion or over-inclusion.[289]

Open meetings laws discriminate among different groups of public officials. These laws regulate legislators but not executive or judicial officials. They regulate only communications among legislators of the same body, not communications between a legislator and an executive branch official of the same government entity. Perhaps most disturbingly, in many states the laws impose burdens on local legislators but exempt state legislators.[290]

It is this latter classification—dividing all legislators into local legislators governed by "sunshine" laws and state legislators who are not—that is most constitutionally problematic and will be discussed here. What basis is there for requiring any two local legislators to have substantive communications about pending matters only via a properly "sunshined" public meeting but exempting two state legislators from any corresponding requirement?

B. Strict Scrutiny

The category of "local legislator" is not one which has previously been recognized as a "suspect class" by the Supreme Court. Such recognized suspect classes normally share such characteristics as a history of discrimination,[291] immutability,[292] and a diminished ability to protect themselves from discrimination through the political process.[293] As a group, "local legislators" cannot plausibly claim a history of official discrimination against them sufficient to trigger heightened review.[294] Membership in this

286. United States v. Virginia, 518 U.S. 515, 532–33 (1996); Craig v. Boren, 429 U.S. 190, 197 (1976).

287. Clark v. Jeter, 486 U.S. 456, 461 (1988).

288. *Id.*

289. Califano v. Boles, 443 U.S. 282, 305 (1979).

290. *E.g.*, DEL. CODE ANN. tit. 29, § 10002(c) (West 2010); 5 ILL. COMP. STAT. ANN. 120/1.02 (West 2010) (carving out an exception for the General Assembly and its subsidiary committees).

291. Frontiero v. Richardson, 411 U.S. 677, 684–88 (1973).

292. *Id.*

293. Lawrence v. Texas, 539 U.S. 558, 580 (2003).

294. There may be other examples of laws which treat state legislators more favorably

class is manifestly mutable: we see its mutability after each election cycle. Compared to an average citizen, local legislators have an influence on the political process that is enhanced, not diminished. From this standpoint, an Equal Protection challenge to open meetings laws which exempt state legislators might be subject merely to rational basis review.

However, there is one argument for heightened review here. Heightened scrutiny is also appropriate under the Equal Protection Clause when the state's classification burdens a fundamental right. The Supreme Court has long held that unequal treatment affecting the right to vote must be evaluated under strict scrutiny. For example, in *Dunn v. Blumstein*, the Supreme Court invalidated Tennessee's durational residency requirement, which required persons to reside in Tennessee for one year, and in the relevant county for three months, in order to vote in a Tennessee county.[295] The Court noted that under Equal Protection, such differing treatment regarding the right to vote required strict scrutiny.[296] The heightened review came not because the affected category of "new residents" was a suspect class, but because Equal Protection demanded strict scrutiny of any differing treatment *regarding the fundamental right to vote*. Similarly, in *Reynolds v. Sims*, the Court struck down Tennessee's state legislative districting scheme as a violation of the Equal Protection Clause's "one person, one vote" principle.[297] Again, strict scrutiny applied because the districts were classifications of voters which affected voting rights.[298] And in *Kramer v. Union Free School District*, the Court applied strict scrutiny to invalidate under Equal Protection a New York law that limited voting in school board elections to persons who owned land in the district or who had children attending school there.[299]

The same heightened equal protection analysis applies for laws treating categories of persons differently regarding First Amendment rights. In *Williams v. Rhodes*, the Court overturned ballot access restrictions for third parties, explaining that classifications burdening First Amendment

---

than local legislators. However, local legislators have not historically been subject to the systematic discrimination relied upon by the Court in recognizing race, alienage, and gender as suspect classes.

295. Dunn v. Blumstein, 405 U.S. 330, 330, 341–44 (1972). Such "durational residency" cases merit strict scrutiny because the classifications involved burden the right of interstate travel. *See Dunn*, 405 U.S. at 343. *See generally* Saenz v. Roe, 526 U.S. 489 (1999) (invalidating durational residency requirement for receipt of welfare payments); Shapiro v. Thompson, 394 U.S. 618 (1969) (invalidating durational residency requirement for receipt of welfare payments).

296. *Dunn*, 405 U.S. at 330, 341–44.

297. 377 U.S. 533, 561–62 (1964).

298. *Id.*

299. Kramer v. Union Free Sch. Dist., 395 U.S. 621, 633 (1969). *But see* Salyer Land Co. v. Tulare Lake Basin Water Storage Dist., 410 U.S. 719, 734–35 (1973) (upholding a law limiting voting in a special-use irrigation district to landowners by applying the rational basis test).

freedoms were subject to strict scrutiny.[300] However, the Court has not been completely consistent on the standard of review in ballot access cases. For example, a plurality of the Court once rejected strict scrutiny in a case involving restrictions the running for other offices by elected officials, such as a ban on judges and other officials running for the legislature while in office, and a "resign to run" provision triggering automatic resignation if an elected official filed for a different office with more than a year left on his term.[301] That plurality distinguished the right of a voter or party to have a candidate of choice on the ballot, which would require strict scrutiny, with the right of a candidate to place his name on the ballot, which would not.[302]

Most relevant for open meetings law purposes, the Court has been more consistent in applying strict scrutiny in equal protection challenges to laws burdening the First Amendment right of *free speech.* In *Police Department of Chicago v. Mosley*, the Court struck down a city ordinance prohibiting picketing near schools because it discriminated between permissible near-school picketing related to labor disputes and forbade the same picketing not related to labor disputes.[303] The Court explained that under Equal Protection analysis, "statutes affecting First Amendment interests [must] be narrowly tailored to their legitimate objectives."[304] Similarly, in *Austin v. Michigan State Chamber of Commerce*, the Supreme Court considered a Massachusetts law which forbade business corporations from making expenditures related to certain referenda, even though such expenditures were allowed for (a) non-corporate organizations with significant treasuries; (b) labor unions; and (c) media corporations.[305] Citing *Mosley*, the Court reiterated that statutory classifications burdening First Amendment rights triggered strict scrutiny.[306]

Based on these precedents, there is a strong argument for applying strict scrutiny in an equal protection challenge to those open meetings laws which burden local legislators but not state legislators. There is definitely a classification between local and state legislators, and that classification burdens the freedom of speech: the right to speak with a colleague about

---

300. Williams v. Rhodes, 393 U.S. 23, 30–32 (1968); *see also* Texas v. White, 415 U.S. 767, 771–72 (1974) (upholding ballot access requirements under the strict scrutiny standard). *But see* Jenness v. Fortson, 403 U.S. 431, 442 (1971) (upholding less restrictive ballot access rules without expressly applying strict scrutiny).

301. Clements v. Fashing, 457 U.S. 957, 963–66 (1982).

302. *Id.* at 966–68.

303. Police Dep't of Chicago v. Mosley, 408 U.S. 92, 95–98 (1972).

304. *Id.* at 101 (citing *Williams*, 393 U.S. at 30–32 (striking down third party ballot access restrictions under Equal Protection and explaining that such analysis required strict scrutiny where First Amendment freedoms are burdened)).

305. *See generally* Austin, v. Mich. State Chamber of Commerce, 494 U.S. 652 (1996).

306. *Id.* at 666 (citing *Mosley*, 408 U.S. at 101). The Court in *Austin* upheld the distinctions under strict scrutiny, noting the governmental interest in preventing the large accumulations of wealth, possible because of the special advantages of the corporate structure, from corrupting the political process.

matters of public concern outside of an advance-noticed public meeting. Assuming strict scrutiny is applied, the distinction between local and state legislators must be narrowly tailored to further a compelling government interest.

There is indeed a compelling governmental interest in ensuring that government business is conducted "in the sunshine," and that the public have access to, and meaningful input toward, the decision-making process of elected legislators. However, it seems a stretch to say that this governmental interest applies to local legislators but not state legislators, or even that the interest is greater with respect to local legislators than state legislators. Presumably, one could argue that because the decisions of local legislators affect citizens' day-to-day lives more, the need for complete citizen access is greater. But this seems a make-weight argument. One could just as easily say that, because state legislators' decisions are more far-reaching, and because state legislators have powers that local legislators do not,[307] it is more imperative to ensure maximum public access to state legislative decision making.

One could not truthfully assert that there are greater opportunities for public access at the state level such that there is a greater *need* at the local level for open meeting laws. Local media tend to cover local legislative action at least as much, if not more, than state legislative action.[308] Further, all things being equal, it is easier for the lay citizen to contact a local legislator than one who is across the state. Again, this analysis, if anything, suggests a greater need for open meeting laws to apply to state legislators. Overall, treating local legislators more strictly than state legislators seems arbitrary and thus unconstitutional, especially inasmuch as the arbitrary discrimination burdens their fundamental right to speak out on matters of public concern.

This conclusion is bolstered by the Supreme Court's recent decision in *Citizens United v. Federal Elections Commission*.[309] In that case, the Supreme Court held that when regulating political speech, the government

---

307. It is a basic principle of state and local law that municipalities and counties are creatures of the state, created by the state, subject to abrogation by the state and possessed of only those powers granted to it by the state. *See, e.g.*, ROMUALDO P. ECLAVEA ET AL., NEW YORK JURISPRUDENCE, CONSTITUTIONAL LAW § 184 (2d ed. 2010); MICHAEL A. PANE, NEW JERSEY PRACTICE SERIES, LOCAL GOVERNMENT LAW § 3:1 (2009). Only the state has sovereignty. Williams v. Eggleston, 170 U.S. 304, 310 (1898) ("A municipal corporation is, so far as its purely municipal relations are concerned, simply an agency of the state for conducting the affairs of government, and as such it is subject to the control of the legislature."). As such, there are innumerable powers which the state has that local governments do not. *Id.* at 309–10.

308. DORIS GRABER, MASS MEDIA & AMERICAN POLITICS 303–04 (7th ed. 2006) (discussing results of surveys showing that local TV stations spend more than half their time on local stories, as opposed to roughly 10% on state stories and roughly 25% on national stories).

309. Citizens United v. FEC, 130 S. Ct. 876, 886 (2010).

could not treat corporations differently from non-corporate entities or individuals.[310] Analyzing the issue at great length, the Court emphasized the need to treat entities and individuals consistently with respect to restrictions on political speech, and it treated arguments for such differing treatment with great skepticism.[311] Although the Court analyzed the case strictly as a First Amendment issue and focused specifically on discrimination between corporate and non-corporate participants in the political process, the case does signal the Court's willingness to intervene to prevent what it sees as arbitrary and disparate treatment burdening the right of individuals to participate in political discussion.[312]

Thus, under strict scrutiny, the discrimination between state and local legislators by some open meetings laws fails for one of two possible reasons. First, it is unlikely that a compelling governmental interest exists for maximizing public access to the deliberations of local legislators which does not equally apply to state legislators. Alternatively, if one characterizes the governmental interest as a more general one in securing public access to legislative deliberations, such open meetings laws are not narrowly tailored to further this interest given that they are substantially under-inclusive.

C. Rational Basis

Even if the above analysis is incorrect, and the standard of review here is rational basis, there is still cause for concern about the constitutionality of sunshine laws which exempt state legislators. A fair-minded observer may

---

310. *Id.* at 903–13.

311. *See generally id.*

312. At the same time, the *Citizens United* case might provide defenders of strict open meetings laws an additional argument. In the recent federal district court case *Asgeirsson v. Abbott*, No. P-09-CV-59, 2011 WL 1157624 (W.D. Tex. Mar. 25, 2011), the Texas Attorney General used the *Citizens United* opinion's validation of campaign disclosure requirements, 130 S.Ct. at 914–916, to argue that disclosure requirements are fundamentally different from outright speech restrictions, and that the Texas open meetings act was more akin to a requirement that public officials disclose the contents of their private communications. *Asgeirsson*, 2011 WL 1157624 at *201–21.

This novel argument may ultimately save strict open meeting acts, but there is significant room for doubt. For one thing, by their plain terms, open meeting acts do more than merely require disclosure of private communications among public officials: they ban the communication in the first place. For another, campaign finance disclosure laws merely require disclosure of the identity of political campaign contributors and the dates and amounts of the contributions, while open meeting acts require the disclosure of the entire content of the communications. By way of example, if public advocacy organizations like the NAACP were required to disclose the content of all communications among their members, they would very likely have viable free speech claims. *Cf.* NAACP v. Alabama, 357 U.S. 449, 460–465 (1958) (stating that compelled disclosure of membership lists compromised not only privacy rights but First Amendment rights of freedom of association and freedom of speech).

be hard-pressed to advance any rational basis for treating local legislators more strictly than state legislators regarding the exercise of their free speech rights.

However, such an equal protection challenge might collapse on certain state law considerations, depending on the particular state's basis for the state-local distinction. In most states where the distinction exists, the state legislature made the distinction in the open meetings law.[313] In a few states, however, the distinction was judicially created based on the dictates of the state constitution.[314] Courts have either decided that the state constitution grants state legislators the authority to meet in secret[315] or that the state constitution deprives the legislature of the power to bind future legislatures in such matters.[316] While a state constitutional requirement does not exempt

313. ALASKA STAT. ANN. § 44.62.310(a) (West 2009); ARK. CODE ANN. § 10-3-305(a) (West 2010); CONN. GEN. STAT. ANN. § 1-225(c) (West 2008) (explicitly exempting the legislature from agenda or notice requirements); GA. CODE ANN. § 50-14-1(e),(f) (West 2010) (explicitly exempting legislature from agenda or notice requirementss); HAW. REV. STAT. § 92-10 (West 2010) (expressly granting authority to the state legislature to set requirements); 5 ILL. COMP. STAT. ANN. 120/1 (West 2010) (exempting the state legislature because it falls outside the statutory definition of "public body"); IND. CODE ANN. § 5-14-1.5-2(a)(1) (West 2007) (not expressly including the General Assembly); KY. REV. STAT. ANN. § 61.810(1)(i) (West 2010) (exempting committees, other than standing committees, from the open meetings law); LA. REV. STAT. ANN. § 42:6.2 (2009) (granting the state legislature express authority to hold closed meetings in a variety of enumerated situations); MISS. CODE ANN. § 25-41-3(a) (West 2010) (placing legislative committees, but not the state legislature itself, within the scope of the statute); N.M. STAT. ANN. § 10-15-2 (West 2009) (carving out a number of open meetings law exceptions relating to the state legislature); OR. REV. STAT. § 192.610(4) (West 2010) (failing to include the state legislature in the statute); S.C. CODE ANN. § 30-4-70(e) (2005) (allowing closed sessions for the General Assembly in certain constitutionally authorized situations); WASH. REV. CODE ANN. § 42.30.020(1) (West 2010) (expressly excluding the state legislature from the open meetings law); WYO. STAT. ANN. § 16-4-402(a)(ii) (West 2010) (expressly excluding the state legislature from the open meetings law); *see* Abood v. League of Women Voters, 743 P.2d 333 (Alaska 1987) (holding that the legislature could exempt itself from the open meetings law); Coggin v. Davey, 211 S.E.2d 708, 710 (Ga. 1975); Sarkes Tarzian, Inc. v. Legislature of State, 104 Nev. 672, 673 (Nev. 1988) (holding that the legislature could make rules exempting it from the open meetings law in some cases).

314. *See* ARK. CODE ANN. § 10-3-305(a) (West 2010); TENN. CODE ANN. § 8-44-102(a) (Supp. 1998); Mayhew v. Wilder, 46 S.W.3d 760, 770 (Tenn. Ct. App. 2001) (holding that the General Assembly does not fall within the definition of "governing body" applicable to the open meetings law due to state constitutional concerns).

315. *See* Ark. Const. of 1874, art. V, § 13 (1874) ("The sessions of each house, and of committees of the whole, shall be open, unless when the business is such as ought to be kept secret."); see also SCHWING, supra note 1, at 131–34.

316. *See Mayhew*, 46 S.W.3d at 770–71. In *Mayhew*, the Tennessee Court of Appeals outlined additional reasons for interpreting the Open Meetings Act as excluding the state legislature. Defining "governing body" as an entity "whose authority may be traced to state, city, or county legislative action," the court reasoned that this excluded the state legislature, whose authority comes from the state constitution. *Id.* The Court also relied on the statutory

a state from the requirements of the federal Equal Protection Clause, such legal considerations might provide a rational basis for the distinction between state and local legislators. Indeed, such rationales might apply more broadly to a number of other states.

Thus, although the different treatment between state and local legislators raises serious equal protection issues, it is difficult to say whether a court would sustain an equal protection challenge. The outcome may depend on whether a reviewing court decides that strict scrutiny was appropriate.

Note that this equal protection analysis is independent of the free speech analysis. Even if the most strict open meetings laws pass First Amendment muster on their own, the laws' inexplicable differentiation between the two sets of legislators may violate the Constitution.

Further, this discussion of under-inclusivity is itself under-inclusive. The above analysis addresses only the most egregious form of under-inclusiveness: the hypocritical decision by some state legislators to exempt themselves from the rigorous requirements imposed upon local legislators. There is no rational basis for applying such requirements to local legislators without also applying them to predecisional consultations by multimember courts, single-headed agencies, or executive officials.[317]

V. POLICY DISCUSSION

A. Policy Problems with the Broader Open Meetings Laws

Because resolution of any constitutional issues turns on the strength of the government interest in broad, strict open meeting laws, consideration of the policies underlying these laws is relevant. And even if the broadest open meetings laws are constitutional, an examination of the policy issues surrounding them is still worthwhile because such laws create serious public policy problems.

1. General: Applying "Transparency" Consistently

The Kansas Supreme Court made a particularly robust First Amendment defense of Kansas's open meetings law in *State* ex rel *Murray v. Palmgren*.[318] In *Palmgren*, litigants asserted an overbreadth challenge to the Kansas statute which barred "a majority of a quorum" of a local legislative body from discussing public business outside of a properly

---

maxim that a statute must expressly bind the state in order to be effective in doing so. *Id.* The first of these two additional rationales might provide an additional rational basis justifying the state-local distinction.

317. *See* Johnson, *supra* note 9, at 15–16.

318. *See generally* State *ex rel.* Murray v. Palmgren, 646 P.2d 1091 (1982).

noticed public meeting.[319] The case dealt primarily with private meetings held by several county commissioners with a representative of a hospital management firm.[320] After confirming that the firm was available to take over management of a public hospital, the commissioners met in a properly noticed public meeting and voted to terminate the existing hospital management firm.[321] Notably, there was no discussion on this matter prior to the vote.[322]

The court rejected the overbreadth challenge in one paragraph which eloquently states the basic policy rationale behind open meetings laws:

> The First Amendment does indeed protect private discussions of governmental affairs among citizens. *Everything changes, however, when a person is elected to public office.* Elected officials are supposed to represent their constituents. In order for those constituents to determine whether this is in fact the case they need to know how their representative has acted on matters of public concern. *Democracy is threatened when public decisions are made in private. Elected officials have no constitutional right to conduct governmental affairs behind closed doors.* Their duty is to inform the electorate, not hide from it.[323]

The court's discussion is a forceful policy argument for having a basic right of public access to government deliberative proceedings. Applied to claims of overbreadth by specific statutes, however, it is arguably superficial both as a policy argument and a legal analysis.

As a policy argument, it may prove too much. If "democracy is threatened when public decisions are made in private," then we should prevent presidents, governors, and mayors from privately conferring with advisors or legislators as part of their decision-making process. After all, in many cases, their deliberations have much more profound impacts on policy than conversations between two legislators. But courts have long acknowledged that executive branch officials have a right to engage in confidential discussions based on the recognition that without a guarantee of confidentiality, they will not receive the same level of candor.[324]

---

319. *Id.* at 1095.

320. *Id.* at 1094–95.

321. *Id.*

322. *Id.*

323. *Id.* at 1099 (emphasis added).

324. United States v. Nixon, 418 U.S. 683 (1974); *see also* Capital Info. Grp. v. State, 923 P.2d 29, 33 (Alaska 1996) (applying executive privilege protections because they encourage "open exchange" of ideas and advice among officials); Wilson v. Brown, 962 A.2d 1122, 1131 (N.J. Super. Ct. App. Div. 2009) (finding the need for free, private consultation and deliberation to be the most important reason for gubernatorial executive privilege).

Similarly, some states elect attorneys general or treasurers.[325] Should they be forbidden from making decisions about whom to prosecute or about which investments to make in private? Courts have also recognized the need for prosecutors to keep their internal deliberations secret to protect the privacy of witnesses and the reputations of targets of investigations.[326] It seems obvious that an elected treasurer might legitimately wish to control the timing of public announcements of investment decisions. Indeed, by allowing elected sheriffs, trustees, and mayors to confer with each other in private, and to confer with selected legislators in private, open meetings laws give a competitive advantage to these officials that is not shared by local legislators. Such officials can assess the legislative body as a whole by having a series of individual conversations with many members, while each legislator must abstain from learning the feelings of, or lobbying, fellow legislators. This under-inclusiveness should make open meetings laws constitutionally suspect.[327] Similarly, an absolute bar on conducting governmental affairs behind closed doors would not protect individuals' privacy when discussing sensitive matters involving personnel disputes, would cause a distinct negotiating disadvantage by mandating public contract negotiations, and would raise any number of legitimate public concerns about confidentiality.

To be sure, open meetings laws do not apply to executive branch officials and many contain exceptions for personnel matters, individual privacy, or contract negotiations. One cannot adequately consider an overbreadth challenge to an open meetings law by reference to over general paeans to government in the sunshine.

As legal analysis, the *Palmgren* opinion may also go too far when it says that "everything changes" when a person is elected to public office, and that elected officials "have no constitutional right to conduct government affairs behind closed doors." The Kansas Supreme Court did not support this statement with actual authority. Indeed, courts have *not* held that there is an unqualified right of public access to governmental deliberations, and they have explicitly acknowledged the authority of governmental actions to deliberate in secret.[328] As explained above,[329] it is by no means clear that elected officials are completely stripped of their First Amendment rights to speak, to whomever they like and whenever they like, about matters of public concern.

---

325. *E.g.*, ALA. CONST. art. V, § 114; ARIZ. CONST. art. V, § 1.

326. *See, e.g.*, Robinson v. Att'y Gen. of U.S., 534 F. Supp. 2d 72, 82–84 (D.D.C. 2008); Blethen Me. Newspapers, Inc. v. State, 871 A.2d 523, 537 (Me. 2005) (refusing, on grounds of privacy interests held by witnesses and victims, to release prosecutor's records absent credible allegation of governmental misconduct).

327. *See supra* Section III.

328. *See supra* discussion accompanying notes 93–103.

329. *See supra* Section III.C.

2. Whether the Broader Version is Truly Necessary to Fulfill Open Meetings Law Goals

The Kansas Supreme Court's articulation of policy rationales for open meetings laws is typical. One commentator wrote that such laws are designed to (1) prevent the self-dealing and corruption of "backroom deals;" (2) allow the public to serve as a check on potential governmental abuse; (3) provide for a more thorough examination of the issues and articulation of policies and rationales; and (4) promote confidence in government.[330] As discussed below, the strictest form of open meetings laws are not necessary to achieve these goals, and in some cases may be counterproductive.

There is no reason to think that the frequency of corrupt backroom deals would flourish were open meetings laws to require half a quorum, or even a full quorum, before triggering the "sunshine" requirement. This is indeed the law in the vast majority of states, and there is no empirical evidence to suggest that such states suffer significantly more corruption than the minority of states which define a "meeting" more broadly.[331] Narrowing the definition of "meeting" in this way need not create a truck-sized loophole. Statutory language could be crafted to forbid legislators from getting around this requirement through a series of small private gatherings among legislators accumulating to a total over a quorum (or half-quorum).[332] This is in line with the general practice of statutes to forbid persons from intentionally or knowingly doing indirectly what cannot be done directly.

Similarly, a narrowing of that sort would still allow the public to serve as a check on government abuse. Recall that after any small gathering of

---

330. Johnson, *supra* note 9, at 17–20. There is no shortage of different formulations of these rationales, including additional rationales. *See, e.g.*, Fenster, *supra* note 9, at 896–902. But the four rationales listed here capture the essence of the arguments.

331. A search of social studies journals uncovered no empirical evidence for a claim of greater corruption among states with more lenient open meetings laws. A search of news articles for the period 2004-2010 among five representative states with a broad definition of "meeting" reaching less than a quorum (Colorado, Connecticut, Illinois, Kansas, and Virginia), plus five representative states using a narrower "quorum rule" (Arizona, California, Michigan, Ohio, and Texas) showed no more reported instances of corruption in the "quorum rule" states. While a comprehensive empirical analysis is outside the scope of this Article, there appears to be no significant evidence that the more speech-friendly quorum rule leads to greater government corruption.

332. *See, e.g.*, Sutter Bay Assocs. v. Cnty. of Sutter, 68 Cal. Rptr. 2d 492, 502–03 (Cal. Ct. App. 1997) (concluding that it is possible for serial meetings to constitute a conspiracy to violate the open meetings law); McComas v. Bd. of Educ. of Fayette Cnty., 475 S.E.2d 280, 289–92 (W.Va. 1996) (listing numerous cases from multiple states holding that individuals could not achieve indirectly what they were forbidden to do directly). For an example of such statutory language barring circumvention of the quorum rule via "in seriatim" meetings, see the Model Open Meetings Law at the end of this Article.

legislators in which they discuss an issue, there will still be a mandatory publicly noticed official meeting. As long as a quorum has not privately met (at once or *in seriatim*), that formal meeting will not simply serve to "rubber-stamp" the predetermined outcome. Publicly open debate and discussion among the remaining members will still be necessary to attain a consensus sufficient for official action, and if the issue is at all controversial, it will still be necessary for legislators to explain the basis for their votes in full view of the public prior to the final vote.

Even in the worst-case scenario, where a series of small gatherings has resulted in a de facto quorum pre-meeting, political reality will require that legislators nonetheless explain their votes on any issue of heightened public interest or wherever there is controversy. If a recalcitrant legislator were to refuse to do so, the actual vote of that legislator will always be made in public.[333] Given all of this, there remain ample avenues for accountability to the public even in a regime that would allow two or three legislators to talk "offline." Indeed, it is precisely upon this set of informal political checks on illicit backroom deals that we have relied regarding the United States Congress for the entire history of our republic.[334]

The above conclusions hold similarly for the addition of exemptions to open meetings laws for topics which merit private discussion. The federal Sunshine Act has a lengthy list of statutory exemptions for personnel matters, trade secrets, information affecting the privacy of individual citizens, law enforcement records, and certain regulatory financial information.[335] Some state laws have similar exemptions.[336] In these jurisdictions, neither public corruption, government abuse, nor public confidence in government is notably worse than in the minority of states with little to no categorical statutory exemptions. Unless the exceptions are worded, applied, or interpreted so broadly as to swallow the rule, effective public access to meetings will be the norm. The public will thus be able to check government abuse and assure itself of the legitimacy of the process.

Assuming that the above analysis is correct, adequate mechanisms exist to prevent corruption and ensure public accountability even in states using the quorum rule or half-quorum rule. The same is true of states with a robust list of exceptions for discussion of sensitive topics. If all that is so, then public confidence in the legislative process is not fatally eroded in such states.

For all the above reasons, it also seems unlikely that legislative discussion, debate, and articulation of policy rationales would become significantly less thorough as a result of narrowing the "meeting" definition. Indeed, there is reason to think the opposite. As noted below, a

333. *See* 5 U.S.C. § 552b(b) (2006).

334. *See* Fenster, *supra* note 9, at 902.

335. *See* 5 U.S.C. § 552b(b).

336. *See supra* Section II.

number of commentators, some citing empirical data,[337] have argued that strict open meetings requirements tend to stifle debate and reduce the quality and detail of collaborative decision making.

3. The Costs of Overly Broad Open Meetings Laws

a. The Main Costs Raised by Commentators

So far, I have focused on whether the benefits of open meetings laws can be achieved with less restrictive rules. This point leads directly to consideration of the significant costs of strict sunshine laws, costs not normally addressed by courts and legislators. A number of commentators have noted that such acts have tended to (1) chill discussion[338] and thus decrease collegial decision making;[339] (2) reduce the actual number of public meetings held;[340] and thus (3) shift authority to staff,[341] or to lobbyists.[342] Similar findings resulted from a comprehensive implementation study[343] commissioned by Congress to assess the effectiveness of the federal Government in the Sunshine Act seven years after its adoption.

Chilling Discussion/Collegial Decision Making. The Supreme Court has recognized that the candor and quality of deliberations can suffer when they are forced to become public. In recognizing a Constitution-based "executive privilege" in *United States v. Nixon*, the Court recognized that "[h]uman experience teaches that those who expect public dissemination of their remarks may well temper candor with a concern for appearances and for their own interests to the detriment of the decision-making process."[344]

The Supreme Court has separately recognized a non-constitutional executive privilege protecting federal government entities from disclosing documents reflecting internal deliberative processes.[345] The Court acknowledged the existence of this privilege in civil discovery in litigation against the federal government as embedded in a statutory exemption for inter-agency or intra-agency memoranda under the federal Freedom of

---

337. *See infra* notes 332–348 and accompanying text (especially references to the Welborn Study and the 1989 Senate Report).

338. *See* KENNETH C. DAVIS & RICHARD J. PIERCE, JR., ADMINISTRATIVE LAW TREATISE § 5.18 (3d ed. 1994); Fenster, *supra* note 9, at 908–09; Johnson, *supra* note 9, at 21–22.

339. *See, e.g.*, DAVIS & PIERCE, *supra* note 338, at 220; Fenster, *supra* note 9, at 908–09; Johnson, *supra* note 9, at 17–20.

340. *See* Johnson, *supra* note 9, at 23–26.

341. *See, e.g., id.* at 26–27.

342. *See* CHARLES H. KOCH, JR., 33 FEDERAL PRACTICE AND PROCEDURE § 8456 (2006).

343. *See* DAVID M. WELBORN ET AL., IMPLEMENTATION AND EFFECTS OF THE FEDERAL GOVERNMENT IN THE SUNSHINE ACT IN 1984: RECOMMENDATIONS AND REPORTS 235–37 (1986) [hereinafter Welborn Study].

344. United States v. Nixon, 418 U.S. 683, 705 (1974).

345. *See* NLRB v. Sears, Roebuck & Co., 421 U.S. 132, 150–51 (1975).

Information Act.[346] Indeed, the Court noted legislative history from that Act which explicitly feared that intra-agency "frank discussion . . . might be inhibited if the discussion were made public," and that the decisions thus made "would be poorer as a result."[347] Therefore, the Court reaffirmed the existence of a non-disclosure privilege available to documents revealing *predecisional* discussion of a policy issue.[348]

Applied to the related context of open meetings laws, such an approach argues for the ability of legislators to confer in private while deliberating (precisely that which is not allowed by open meetings laws), relying on the public disclosure of the actual decision itself made at a public meeting to ensure adequate public oversight. In effect, the statutory requirement that properly noticed public meetings precede actual action ensures the disclosure of *post-decisional* discussion. The public is informed of which elected official decided what and why. Sufficiently great public outcry can then force reconsideration of the decision, or future decisions of that kind can be prevented by voting the officials out of office. Such an approach can also be reconciled with open meetings laws that adopt a "quorum rule." Once a quorum has met to decide something, the decision is effectively made, and all further discussions are de facto post decisional.

Commentators agree with the Supreme Court that private consultation can enhance the decision-making process:

> Closed deliberations enable policymakers to make more thoughtful consideration of the available information and the relative advantages of alternatives, to engage in more fulsome and substantive debate over the most popular and unpopular alternatives regarding even the most passionate public issues, and to bargain openly in order to reach a widely acceptable and optimal result, without the inevitable pressure that accompanies public scrutiny.[349]

Many of these same advantages support our universal practice of having multi-judge panels and juries deliberate in private. One cannot imagine a state appellate or supreme court, let alone the United States Supreme Court, being required to deliberate controversial decisions in public. Yet many of these decisions have a much more wide-ranging and profound impact on the lives of the citizenry than the type of local ordinance covered by open meetings laws. Similarly, juries make important decisions, even life-or-death decisions, yet the privacy of juror deliberations is considered so sacrosanct that attempts to pierce the veil of secrecy in the most trivial of jury cases can lead to criminal punishment. While not completely analogous to legislative deliberation, these examples do

346. *Id.* at 149 (discussing 5 U.S.C. § 552b(b)(5) (2006)).

347. *Id.* at 150 (internal quotation omitted).

348. *Id.* at 151–52. The pre-decision/post-decision distinction has been echoed by commentators.

349. Fenster, *supra* note 9, 908; *see also* Johnson, *supra* note 9, at 26–29.

illustrate society's recognition that as a practical matter, private deliberation is appropriate and necessary for proper decision making. So too with legislators: private deliberation can lead to greater candor and more nuanced outcomes.

The cramped restrictions of modern open meetings laws thus have a predictable effect:

> Anecdotal complaints about open meeting laws suggest that agencies subject to these laws hold fewer meetings; engage in a constrained, less-informed dialogue when they meet; are vulnerable to greater domination by those who possess greater communications skills and self-confidence, no matter the quality of their ideas; and lose the potential for informal, creative debate that chance or planned meetings outside of the public eye enable.[350]

This stifling of debate is aggravated where the subject matter is sensitive and the relevant open meetings law admits of few or no subject matter exceptions. For example, suppose a legislative body needs to make an appointment to some government position to fill a vacancy. Members may wish to candidly discuss the pros and cons of various candidates for the vacancy, including reviewing negative information on a candidate's background that may potentially embarrass the candidate. Without an appropriate exception for personnel matters, matters that may infringe on a citizen's privacy, or the like, many legislators might simply decline to raise the issue, thus depriving the body of relevant information and weakening the decision-making process.[351]

Moreover, even so simple a thing as co-sponsorship becomes problematic when such laws prevent any two legislators from conferring privately. A legislator drafting a bill *may not* ask colleagues to co-sponsor the bill prior to its public release. Once it is formally introduced, of course, a legislator may publicly ask for co-sponsors. But some legislators may be reluctant to introduce a controversial bill in the first place unless they know that key colleagues—either those of the same party, or perhaps of the opposite party—will co-sponsor with them. Democracy is furthered, not subverted, by allowing a sponsor to seek such early support in an off-the-record discussion prior to the formal introduction of the bill.

Fewer Meetings. A 1989 Senate Report studying the Sunshine Act's effects on the federal government showed a 31% decline in all federal agency meetings held between 1980 and 1984, based on a survey of fifty-

---

350. Fenster, *supra* note 9, at 909.

351. There are still other arguments for the proposition that overly rigid public access rules weaken legislative and other governmental output. *See id.* at 909–10 ("Just as creativity and innovation in the sciences and arts are adversely affected by a legal regime that under-protects intellectual property, so the amount of information produced by government and the quality of its decision making are harmed when disclosure requirements become too rigorous.").

nine federal agencies.[352] A few years earlier, the Welborn Study found that, after the Sunshine Act's passage, federal agencies engaged in greater use of "notation voting," decision making "on the papers" without actual meetings.[353] One commentator has suggested that open meetings laws encourage greater use of the related device of the "consent agenda," where unanimously supported items are bunched together and resolved without discussion through a single vote.[354]

Reliance on Staff. The Welborn Study found that the number of staff meetings increased after adoption of the Federal Sunshine Act.[355] Such meetings were more common particularly right before scheduled open meetings.[356]

Hardly surprising, such a result suggests that agency members asked staff to meet to hash out issues prior to formal meetings. While an understandable instinct, it naturally tends to place more discretion in the hands of staff and less in the hands of the agency members or legislators accountable to the public.

The shift of power to staff is an intuitive result, one entirely in accord with the author's own experience as a local legislator. The more complex or controversial an issue, the greater the impulse of a legislator to confer with colleagues about it in private. Since a legislator cannot confer privately with a fellow decision maker, the legislator naturally turns to staff for guidance, even more than the legislator otherwise might. Further, unlike the legislator, staff members are allowed to consult with multiple legislators and get an overall view of where the legislative body is on a given issue. This information advantage enhances staff members' ability to frame the debate and guide the outcome, and places them in a heightened role as mediator between competing positions of individual legislators. The result is a transfer of power from those elected by the people to unelected bureaucrats.

A similar dynamic is at work with respect to lobbyists and executive branch officials. When complicated or controversial issues are taken up by a legislative body, discussion often continues over a series of formal public meetings. In resolving any policy impasses, it is crucial to know where each legislator stands on the issue and what compromises each is prepared to accept. A lobbyist or executive branch official is free to contact each

352. ROGELIA GARCIA, CONG. RESEARCH SERV. REPORTS, *Government in the Sunshine: Public Access to Meetings Held Under the Government in the Sunshine Act, 1979–1984*, *in* GOVERNMENT IN THE SUNSHINE ACT: HISTORY AND RECENT ISSUES, S. REP. No. 101-54, at 61, 63 (1989).

353. Welborn Study, *supra* note 343, at 236–39.

354. Johnson, *supra* note 9, at 25–26. This assertion may overstate the chilling effect of sunshine laws. Many local legislative bodies routinely use the consent agenda as a time-saving tactic as part of their regular rules of order. *See, e.g.*, Shelby County, Tenn., Permanent Record of Order of the Board of County Commissioners (2010). Such routine usage may be unaffected by the strictness or laxity of the applicable open meetings law.

355. Welborn Study, *supra* note 343, at 223.

356. *Id.*

legislator and discover exactly what that legislator's position is at any given phase of the process. A lobbyist thereby learns which compromises are feasible and which are unrealistic. This gives the lobbyist an enormous tactical advantage over an individual legislator, who is barred by law from finding out where any colleague stands. In this way, broad sunshine laws transfer power from the legislature to the executive and from elected legislators to unelected lobbyists.

Such a transfer exacerbates the disadvantage faced by local legislators, almost all of whom are part-time officials. Most such legislators have full-time "day jobs" they use to support themselves and are thus limited in the amount of time and study they can devote to complex policy issues.[357] Therefore, they are often forced to rely on the greater expertise of full-time staff, executive branch officials, and lobbyists when making up their minds. By isolating each legislator from fellow legislators outside the limited venue of formal public meetings, open meetings laws make this power dynamic even more lopsided. *Quaere* whether this truly enhances the democratic process.

b. Other Costs

In their broadest form, open meetings laws create still more problems. These problems have not been discussed in detail by commentators.

Reduces Efficiency. Obviously, the requirement of a publicly noticed meeting for any discussion between any two legislators slows the resolution of legislative issues. While it is generally understood that democracy is necessarily an inefficient process,[358] taken to this extreme, sunshine laws can cause significant problems for the part-time local legislator. If the legislative body is taking up a complicated issue requiring lengthy legislation, there may simply not be enough time to work out all the details during formal meetings, which often involve lengthy agendas and members of the public and staff waiting for particular items to be heard so they can leave.

An obvious time-saving solution would be for key members of the legislature to meet informally to hash out a tentative proposal which would then be discussed openly at the next regularly scheduled meeting. Deprived of this sensible solution by the strictest of the open meetings laws, legislators are faced with three bad choices: (1) repeatedly postponing decisions while the details get worked out through a series of successive regularly scheduled meetings, usually at two-week intervals; (2) scheduling a special meeting to work on the issue, despite the crowded and conflicting

357. *See* Pam Squyres, *Legislators' Day Jobs*, MOTHER JONES, May 22, 2000, http://motherjones.com/politics/2000/05/legislators-day-jobs.

358. ADAM PRZEWORSKI ET AL., DEMOCRACY AND DEVELOPMENT: POLITICAL INSTITUTIONS AND WELL-BEING IN THE WORLD, 1950-1990 14–15 (2000).

schedules of part-time legislators with "day jobs;" or (3) taking action based on incomplete debate and discussion.

Prevents Compromise. Another problem with broad open meetings laws is that they make legislative compromises on divisive issues more difficult. When parties are locked in a bitter impasse, it is often useful for one member to privately reach across the aisle and float a potential compromise.

Doing so in public entails great risk. The other side may decide to yield political gain by publicly rebuffing the suggestion, playing to its base by loudly decrying any "sell-out" and embarrassing the member who made the suggestion. Or the other side may wish to negotiate but feel constrained from doing so publicly by pressure from interest groups or hard-liners on its own side.

The risk is even greater when, as is often the case, the compromise is multilateral. A promoter of a compromise must often speak in hypotheticals, asking A if A would yield on Issue 1 if the promoter could get B to yield to A on Issue 2; the promoter might continue that if, and only if, that were to take place, the promoter personally would be willing to yield on Issue 3. Such multi-party negotiations are inherently delicate and must often be carried out in stages. In Stage One, it might be politically risky, and fatally so, for A to publicly give conditional, hypothetical assent without yet knowing whether the other parties will be willing to go along. This chilling effect can abort the incipient compromise.

Tacitly acknowledging this reality, media members often praise members of Congress for privately "working across the aisle" to broker compromise and break gridlock.[359] It is not reading too much into such praise to see a realization that such delicate negotiations might break down if the participants were forced to negotiate in public. Yet many of these same media commentators would vehemently condemn any attempt to narrow open meetings laws applicable to local legislators, calling such efforts an attempt to return to the smoke-filled room.[360]

Forces Inappropriate Disclosure of Sensitive Information. As noted above, absent an appropriate sunshine law exception, a legislator may decide not to raise a sensitive matter for fear of embarrassing an individual or harming that individual's reputation. Alternatively, the legislator may feel obligated to raise the matter in public, doing otherwise unnecessary damage to the individual. Indeed, the prospect of raking over a job candidate's record in public may dissuade some qualified candidates from applying for such positions, lest they endure the harsh glare of public

359. *See Primary Choices: John McCain*, N.Y. TIMES, Jan. 25, 2008, at A24, *available at* http://www.nytimes.com/2008/01/25/opinion/25fri2.html.

360. *See Beef Up the Open Meetings Law*, N.Y. TIMES, June 29, 1994, at A22, *available at* http://www.nytimes.com/1994/06/29/opinion/beef-up-the-open-meetings-law.html. Of course, not every media discussion of sunshine laws opposes such reforms.

scrutiny.[361] Such a result naturally harms both the candidates and the public institution searching for them.

Similarly, a legislator may feel politically obligated to discuss the city or county's potential "bottom line" in ongoing labor talks, or in a negotiation for the sale of land to, or purchase of goods or services from, a private entity. Doing so may substantially weaken the city or county's bargaining position. Such dilemmas pit the legislator's obligation to protect the government's financial interest against the legislator's obligation to engage in full consideration and discussion in accordance with applicable law.

Rewards the Scofflaw and Punishes the Scrupulous. Given the many disadvantages to the legislator entailed in strict adherence to broad open meetings laws, it should come as no surprise to learn that such laws are often honored in the breach. Reported instances of substantial violations are not uncommon.[362]

Yet another pressure to violate strict sunshine laws comes from the competitive nature of legislative politics. Legislators often compete with one another, not only over competing policy visions, but over issues such as budgetary resources, credit for policy initiatives, and bragging rights over legislative victories. The legislators who know what their colleagues are thinking at all times—including, and especially, prior to regularly scheduled public meetings—have a distinct comparative advantage. These are the legislators who end up advancing their legislative agendas, brokering deals, and earning reputations for "getting things done" and being "the guy to see" on Issue X. This, in turn, leads to prestige and influence. The legislators who most scrupulously honor the sunshine law, and are thus the most in the dark about colleagues' positions until the formal debate, are less likely to achieve their policy goals, less likely to broker deals, and generally will have a lower profile.

While it may always be the case that "cheaters" have an unfair advantage over those who play fair, at least until the cheaters are caught, the problem is exacerbated where a rule widely seen as an unrealistic technicality is routinely broken by a wide variety of actors. This, sadly, is almost certainly the case regarding the broadest open meetings laws.

Breeds Contempt for the Law. This last observation illustrates a related but distinct, pernicious byproduct of overbroad sunshine laws. By

---

361. *Cf.* Fenster, *supra* note 9, at 908 n.104 (2006) (noting that the pool of applicants for high level administrator jobs at public universities has been narrowed by the application of open meetings laws to such job searches) (citing Nick Estes, *State University Presidential Searches: Law and Practice*, 26 J.C. & U.L. 485, 502–08 (2000)).

362. *See, e.g.*, State *ex rel* Murray v. Palmgren, 646 P.2d 1091, 1101 (1982); *Readers Cheer Tenn. Newspaper's Open-Meetings Lawsuit*, FIRST AMENDMENT CENTER (Mar. 20, 2007) http://www.firstamendmentcenter.org/news.aspx?id=18307; *see also* Michele Bush Kimball, *Law Enforcement Records Custodians' Decision-Making Behaviors in Response to Florida's Public Records Law*, 8 COMM. L. & POL'Y 313, 314–15 (2003) (discussing pervasive noncompliance by state and local government agencies)

creating a regime in which violation of the rules is commonplace, such laws breed contempt for the law. Political actors in such regimes routinely joke about the open meetings law. Each actor feels free to craft his or her own "exceptions:" situations where the actor unilaterally decides that a certain violation of the open meetings law is merely technical in nature and not worth worrying about. The practices vary from person to person, creating confusion among legislators regarding both what the law is as a nominal matter and the actual state of compliance as a realistic matter.

The situation is not different from any unrealistic "zero tolerance" law. If a high school student knows that the punishment for being caught with a pseudophedrine tablet is essentially the same as for being caught with a marijuana joint, that student tends to take less seriously both the dangers of marijuana and the authority of the school. This insouciance transfers over to other rules, leading to an epidemic of scofflaw behavior.[363]

B. Model Open Meetings Law

A proposed Model Open Meetings Law is set out below. It covers legislative bodies and their subsidiary agencies but not those agencies with merely advisory or ceremonial duties. It explicitly requires that state and local bodies be treated alike. Regarding the crucial definition of "meeting," the Model Law adopts the "quorum rule" used by a majority of states and compiles certain typical categorical exceptions for topics that may appropriately be treated as confidential. In addition to personnel matters, matters affecting individual privacy, and ongoing financial negotiations, these exceptions also explicitly allow a bill sponsor to seek co-sponsors. Since discussions of such topics are not "meetings," they are not covered by the Model Law, and individual members amounting to less than a quorum can have informal discussions about these topics.[364] The Model Law allows for retreats by the covered government entity and echoes the exception for fact-finding meetings present in a number of states' open meetings laws. Additionally, the Model Law provides a defined procedure for closing a formal meeting. The Model Law is, by design, simple and short.

As is typical, the enforcement mechanism is a private lawsuit by an interested party. Because criminal liability entails a substantial likelihood of

---

363. Nekima Levy-Pounds, *Can These Bones Live? A Look at the Impacts of the War on Drugs on Poor African-American Children and Families*, 7 HASTINGS RACE & POVERTY L.J. 353, 371 (2010) (discussing the tendency of zero tolerance policies to lead to juvenile delinquency).

364. This is consistent with another Model Open Meetings Law drafted by commentators. *See* Little & Tompkins, *supra* note 9, at 485 (setting out a model law with the proviso that "[n]othing herein shall make illegal informal discussions, either in person or telephonically, between members of public bodies for the purpose of obtaining facts and opinions provided that there is no intention of violating [the law]"), *quoted in* St. Cloud Newspapers, Inc. v. District 742 Cmty. Schs., 332 N.W.2d 1, 8–9 (Minn. 1983) (Simonett, J., concurring in part and dissenting in part).

chilling free speech, the remedies do not include criminal sanctions.

However, the remedies do include civil penalties for individual legislators and members of boards and commissions, but only after a showing of willful misconduct: where one member conspired with others to violate the Model Law, such as where two members of a government body agree to hold a series of *in seriatim* meetings or telephone calls to achieve a quorum cumulatively. Because this is a civil penalty imposed on an individual, the heightened proof standard of clear and convincing evidence is used. By providing for the shifting of costs and attorney fees, the Model Law also seeks to encourage vindication of the rights provided by "private attorneys general." On the other hand, to discourage frivolous and politically motivated lawsuits, the law allows for costs and attorney fees to be assessed against the plaintiff based on a finding of a frivolous claim.

Model Open Meetings Law

1. General. This Act applies to all legislative bodies within this State and all multimember boards, commissions, and agencies appointed by such a legislative body that have the ability to issue rules or decisions which, if left undisturbed, are legally binding. It applies equally in all respects to state and local bodies.

2. Requirement of Open Meetings. All meetings of covered government entities must be open to the public and properly noticed to the public at least 48 hours in advance. Notice shall include the name of the covered body, the time and place of the meeting, a copy of the agenda, and a statement of whether minutes, a transcript, or a recording of the meeting will be made available. Notice shall be accomplished through, at a minimum, placement of a written notice on a designated public bulletin board and on the applicable state, county or city website, if any. The covered government entity may devise additional methods of notice.

3. "Meeting" Defined.

(a) General Definition. For purposes of this Act, a "meeting" is any communication, whether in person, in writing, or through some form of electronic communication, among a quorum of the relevant government entity to the extent such communication involves deliberation toward an official decision by that government entity. A member of a covered government entity may not intentionally circumvent this provision by participating, directly or indirectly, in a series of communications among other members less than a quorum which, taken together, involve a number of such members equal to or greater than a quorum.

(b) Exceptions. The term "meeting" shall not include:

(1) Fact-finding trips, site inspections, or the like;

(2) Retreats sponsored by the government entity, provided that such retreats occur no more frequently than quarterly; or

(3) Discussions of:

(i) personnel decisions, including appointments to fill vacancies in elected or appointed governmental positions;
(ii) trade secrets, confidential intellectual property, or other commercial proprietary information, including, but not limited to, information which, if disclosed by an employee or competitor, would normally give rise to civil liability;
(iii) financial, medical, or other sensitive information concerning a private business or individual that would disturb personal privacy, including, but not limited to, information which, if disclosed by a private party, would normally give rise to tort liability for invasion of privacy;
(iv)then-pending litigation, administrative adjudicatory proceedings, or official investigations into violations of law, ordinance, or regulation;
(v) any information which an applicable statute requires or permits to be held confidential;
(vi) then-pending commercial negotiations between the government entity and another individual or entity, public or private; or
(vii) a potential sponsor's request that a colleague co-sponsor draft legislation.

4. Closed Meetings. A formal meeting of a covered government entity may be ordered closed to the public by a majority vote of the government entity, provided that the general counsel of the entity, or of the legislative body appointing it, or some other qualified consulting attorney, advises that one of the exceptions of Section 3(b) applies. In making this determination, a presumption in favor of open meetings shall apply. Discussion at the closed meeting must be kept pertinent to the matters triggering such exception. No final action can be taken in a closed meeting.

5. Remedies. Any resident of the political jurisdiction in or for which the covered government entity acts may file an action in a court of competent jurisdiction to enforce this law. The court may order, as appropriate:

(a) an injunction ordering an upcoming meeting open to the public;
(b) an injunction nullifying an action taken in violation of this Act, which action may be reinstated by a subsequent vote of the covered government entity done in compliance with this Act;
(c) an injunction against future violations of the Act;
(d) costs and attorney fees against the covered jurisdiction after a finding of a violation of the Act, or against the plaintiff after a finding that his or her claim was frivolous;

(e) civil penalties against the covered government entity after a finding that it took a frivolous position in the litigation;
(f) civil penalties against an individual member of the government entity, after a finding by clear and convincing evidence that such member willfully conspired with others to violate the act;
(g) such other relief as the court in the exercise of reasonable discretion deems appropriate and consistent with the provisions of this law.

VI. CONCLUSION

Open meetings laws are content-based restrictions on political speech that deserve strict scrutiny. Where they reach down to regulate individual conversations between two legislators who may wish to confer in private, there is a serious doubt as to whether they are narrowly tailored. Similar doubt exists where such laws contain no exceptions to protect individual privacy, to allow legislative clients to confer confidentially with counsel, or to permit local government agencies to negotiate with outside vendors in private. Even under the more forgiving constitutional standard used for content-neutral regulations of speech, these stricter laws may be fatally under-inclusive or over-inclusive or fail to provide ample alternative channels for deliberation.

It is telling indeed that many state legislatures exempt themselves from the strictest of the open meeting requirements they impose on local government entities. Their tacit acknowledgment of the difficulties involved in banning all private deliberation is understandable, but it is also in tension with equal protection principles.

Discussions of open meetings policies inevitably turn to Justice Brandeis' famous maxim that "[sunlight] is said to be the best disinfectant."[365] Comparing a right of public access to sunshine is a powerful metaphor, but, like most metaphors, it can work in multiple directions. Sunlight cannot really disinfect, but overexposure can cause sunburn, skin cancer, and heat exhaustion. In a similar manner, champions of good government certainly should insist that the public be informed of all important government decisions while they are made and that formal public meetings not be sham affairs in which backroom deals are rubber-stamped. But that does not mean that legal sanctions are appropriate every time a Republican legislator takes a Democratic counterpart by the elbow and says, "Let's go get some coffee and see if we can work out a compromise." Nor does it mean that a school board must do live web

365. LOUIS BRANDEIS, OTHER PEOPLE'S MONEY AND HOW THE BANKERS USE IT 92 (1914). The statement was made not in the context of open meetings laws or public access to government decision making but rather activity by private industry. Specifically, the statement refers to proposed regulations requiring disclosure of financial information to shareholders and the public by banks and institutional investors. *Id.* Nonetheless, it is quoted commonly as a call for open government.

streaming when it considers a grievance from a principal accused of sexually harassing a minor.

Open government reform is thus, itself, in need of some reform. The most appropriate vehicle for such reform would be state legislation in line with the Model Open Meetings Law set out above. However, self-interested opposition from media makes such legislative reform difficult to achieve. Arguing for more secrecy in government is a tough sell to a distracted public under the best of circumstances; add inflammatory editorials about "smoke-filled rooms," and such reform may be impossible. Absent such reform, courts may see more challenges like *Rangra*. One way or another, hopefully local legislators may eventually find some relief from the sunlight's glare.

VII. APPENDIX

Subject-Matter Exceptions to Open Meetings Law Requirements by State

This table reflects the topics which are not covered by state open meetings laws. An "X" indicates that the relevant state's open meetings law does not apply to discussions of the topic described in the column.

State	Protection of individual privacy	Personnel matters	Student disciplinary hearing	Attorney consultation/ litigation	Public safety	Certain law enforcement matters	Investigative proceedings	Acquisition or sale of real property	Preliminary negotiations involving trade or commerce/ To protect confidential data or trade secrets	Labor negotiations	Certain proceedings of a quasi-judicial body
Alabama	X	X	X	X	X	X	X	X	X	X	X
Alaska	X										X
Arizona		X		X				X		X	
Arkansas		X			X						
California	X	X		X	X			X			
Colorado		X	X	X	X			X		X	
Connecticut		X		X	X			X		X	
Delaware		X	X	X			X	X		X	
Florida				X	X				X		
Georgia		X		X		X	X	X			
Hawaii		X		X	X		X	X		X	
Idaho		X	X	X				X	X	X	
Illinois		X	X	X	X	X	X	X		X	X
Indiana		X	X	X	X			X	X	X	
Iowa		X	X	X		X		X			
Kansas		X	X	X	X			X	X	X	X
Kentucky		X	X	X				X			X
Louisiana	X		X	X	X		X			X	
Maine		X	X	X				X		X	
Maryland	X	X		X	X		X	X	X	X	
Massachusetts	X	X		X	X		X	X		X	
Michigan		X	X	X				X		X	
Minnesota				X	X			X		X	X
Mississippi		X	X	X	X	X	X	X	X		
Missouri	X	X	X	X	X			X	X	X	
Montana	X			X							
Nebraska		X		X	X		X	X		X	
Nevada	X										
New Hampshire	X	X	X	X	X			X		X	
New Jersey	X	X		X	X		X	X		X	
New Mexico	X	X		X				X	X	X	
New York		X		X	X	X	X	X		X	X
North Carolina		X		X	X		X	X	X	X	
North Dakota				X							
Ohio		X		X	X		X	X		X	
Oklahoma		X	X	X	X			X		X	
Oregon		X		X	X			X	X		
Pennsylvania		X		X				X		X	
Rhode Island	X		X	X	X		X	X	X	X	
South Carolina		X	X	X	X		X	X	X		
South Dakota		X	X	X					X	X	
Tennessee									X		
Texas		X	X	X	X			X	X		
Utah	X			X	X		X	X		X	
Vermont		X	X	X	X			X		X	X
Virginia	X	X	X	X	X	X		X	X		
Washington		X		X	X			X	X		
West Virginia	X	X	X		X		X	X			X
Wisconsin		X		X	X			X	X		X
Wyoming		X	X	X	X			X		X	

State	Caucuses	Courts	Voluntary membership associations with no legislative or executive functions	Matters that would have an adverse effect upon the finances of the public entity	To consider or receive information classified as confidential by law	Juries	Parole boards	Certain meetings of members of public hospitals	Staff meetings or other gatherings of employees of a public entity	Preparation and grading of examination materials/ Application for professional license	Certain administrative proceedings
Alabama	X	X	X								
Alaska		X	X	X	X	X	X	X	X		
Arizona	X	X			X		X				
Arkansas						X				X	
California					X		X			X	X
Colorado					X			X			
Connecticut	X				X				X		
Delaware		X			X	X	X				
Florida											
Georgia					X		X	X			
Hawaii					X					X	
Idaho					X		X				
Illinois			X		X		X	X			
Indiana	X		X		X					X	
Iowa					X					X	X
Kansas							X				
Kentucky					X	X	X				
Louisiana		X								X	
Maine					X					X	
Maryland		X			X	X		X		X	
Massachusetts											
Michigan	X				X						
Minnesota					X						
Mississippi		X			X	X	X	X		X	
Missouri					X					X	
Montana		X									
Nebraska		X								X	
Nevada											
New Hampshire	X						X				
New Jersey		X			X	X	X			X	
New Mexico		X						X		X	X
New York	X	X							X	X	
North Carolina					X			X			
North Dakota	X				X						
Ohio					X	X	X	X		X	
Oklahoma		X			X						X
Oregon		X			X	X	X	X		X	
Pennsylvania	X				X						
Rhode Island	X										
South Carolina								X			
South Dakota											
Tennessee											
Texas					X		X	X		X	
Utah											
Vermont		X			X		X				
Virginia	X				X	X	X			X	
Washington											
West Virginia	X				X					X	
Wisconsin		X					X				
Wyoming					X		X			X	

State	To consider the conferring of honorary degrees or awards	Certain matters related to gifts, donations, and bequests	Investment of public funds	Certain tax matters	Public insurance agencies	Child death review team	Ethics commission	Certain proceedings of municipal utilities	Cabinets or executive councils	Judicial nominating commission	Certain workers' comp proceedings
Alabama											
Alaska										X	
Arizona											
Arkansas											
California	X	X	X		X						
Colorado	X										
Connecticut											
Delaware		X		X							
Florida				X							
Georgia			X								
Hawaii		X									
Idaho				X	X			X			
Illinois			X		X	X	X	X			
Indiana											
Iowa											
Kansas											X
Kentucky									X	X	
Louisiana											X
Maine											
Maryland			X		X				X	X	
Massachusetts											
Michigan	X				X						X
Minnesota											
Mississippi				X							X
Missouri											
Montana											
Nebraska									X		
Nevada											
New Hampshire											
New Jersey											
New Mexico											
New York	X		X								
North Carolina	X										
North Dakota											
Ohio				X		X					
Oklahoma											
Oregon			X								X
Pennsylvania							X				
Rhode Island		X	X								
South Carolina			X								
South Dakota											
Tennessee											
Texas		X	X		X						
Utah											
Vermont											
Virginia	X	X	X			X					
Washington			X								
West Virginia	X										
Wisconsin			X		X		X				X
Wyoming		X									

The exceptions listed above are derived from the following statutory provisions:

Alabama: ALA. CODE § 36-25A-2(4) (2010); ALA. CODE § 36-25A-7(a) (2010).

Alaska: ALASKA STAT. § 44.62.310(c)–(d) (2009).

Arizona: ARIZ. REV. STAT. ANN. § 38-431.03(A) (2010); ARIZ. REV. STAT. ANN. § 38-431.08(A) (2010).

Arkansas: ARK. CODE ANN. § 25-19-103(4) (West 2010); ARK. CODE ANN. § 25-19-106(c) (West 2010).

California: CAL. GOV'T CODE § 11126(c), (e)(1) (West 2009); CAL. GOV'T CODE § 54956.7–54957.10 (West 2009).

Colorado: COLO. REV. STAT. ANN. § 24-6-402(3)(a) (West 2010); COLO. REV. STAT. ANN. § 24-6-402(4)(a) (West 2010).

Connecticut: CONN. GEN. STAT. ANN. §1-200(2), (6) (West 2010).

Delaware: DEL. CODE ANN. tit. 29, § 10004(b), (h) (West 2010).

Florida: FLA. STAT. ANN. § 286.011(3), (8) (West 2010).

Georgia: GA. CODE ANN. § 50-14-3 (West 2010).

Hawaii: HAW. REV. STAT. § 92-5(a) (West 2010).

Idaho: IDAHO CODE ANN. § 67-2345 (1) (West 2010).

Illinois: 5 ILL. COMP. STAT. 120/1-02 (2008); 5 ILL. COMP. STAT. 120/2(c) (2008).

Indiana: IND. CODE § 5-14-1.5-2(c) (2007); IND. CODE § 5-14-1.5-6.1(b) (2007).

Iowa: IOWA CODE ANN. § 21.5 (West 2010).

Kansas: KAN. STAT. ANN. § 75-4318(f) (2010); KAN. STAT. ANN. § 75-4319(b) (2010).

Kentucky: KY. REV. STAT. ANN. § 61.810(1) (West 2010).

Louisiana: LA. REV. STAT. ANN. § 42:6.1(A)–(B) (2010).

Maine: ME. REV. STAT. ANN. tit. 1, § 405(6) (2010).

Maryland: MD. CODE ANN., STATE GOV'T § 10-502(h)(3) (West 2010); MD. CODE ANN., STATE GOV'T § 10-508(a) (West 2010).

Massachusetts: MASS. GEN. LAWS ANN. ch. 30A, § 11A 1/2 (West 2010); MASS. GEN. LAWS ANN. ch. 39, § 23B (West 2010).

Michigan: MICH. COMP. LAWS ANN. § 15.263(7) (West 2010); MICH. COMP. LAWS ANN. § 15.268 (West 2010).

Minnesota: MINN. STAT. § 13D.01(2) (2009); MINN. STAT. § 13D.03 (2009); Minn. Stat. § 13D.05 (2009).

Mississippi: MISS. CODE ANN. § 25-41-3(a) (West 2010); MISS. CODE ANN. § 25-41-7(4) (West 2010).

Missouri: MO. ANN. STAT. § 610.021 (West 2010).

Montana: MONT. CODE ANN. § 2-3-203(3)–(5) (2010).

Nebraska: NEB. REV. STAT. § 82-1409(1)(b) (2006); NEB. REV. STAT. § 82-1410(1) (2006).

Nevada: NEV. REV. STAT. ANN. § 241.030(1) (West 2010).

New Hampshire: N.H. REV. STAT. ANN. § 91-A:2(I) (2010); N.H. REV. STAT. ANN. § 91-A:3(II) (2010).

New Jersey: N.J. STAT. ANN. § 10:4-8(a) (West 2010).

New Mexico: N.M. STAT. ANN. § 10-15-1(A), (H) (West 2010).

New York: N.Y. PUB. OFF. LAW § 105(1) (Consol. 2009); N.Y. PUB. OFF. LAW § 108 (Consol. 2009).

North Carolina: N.C. GEN. STAT. ANN. § 143-318.10(c) (West 2010); N.C. GEN. STAT. ANN. § 143-318.11(a) (West 2010).

North Dakota: N.D. CENT. CODE § 44-04-19.2(1)–(2) (2007); N.D. CENT. CODE § 44-04-19.3 (2007).

Ohio: OHIO REV. CODE ANN. §121.22(D)–(G) (West 2010).

Oklahoma: OKLA. STAT. ANN. tit. 25, 304(1) (West 2010); OKLA. STAT. ANN. tit. 25, 307(B) (West 2010).

Oregon: OR. REV. STAT. § 192.660(2) (2009); OR. REV. STAT. § 192.690(1) (2009).

Pennsylvania: 65 PA. CONS. STAT. ANN. § 703 (West 2010); 65 PA. CONS. STAT. ANN. § 707(c) (West 2010).

Rhode Island: R.I. GEN LAWS § 42-46-2(c) (2007); R.I. GEN LAWS § 42-46-5(a) (2007).

South Carolina: S.C. CODE ANN. § 30-4-20(a) (2010); S.C. CODE ANN. § 30-4-70(a) (2010).

South Dakota: S.D. CODIFIED LAWS § 1-25-2 (2009).

Tennessee: TENN. CODE ANN. § 8-44-102(b) (2010); TENN. CODE ANN. § 50-3-2013(c)(1) (2010).

Texas: TEX. GOV'T CODE ANN. § 551.071–088 (West 2007).

Utah: UTAH CODE ANN. § 52-4-205(1) (West 2010).

Vermont: VT. STAT. ANN. tit. 1, § 312(e) (2010); VT. STAT. ANN. tit. 1, § 313(a) (2010).

Virginia: VA. CODE ANN. § 2.2-3703(A) (West 2010); VA. CODE ANN. § 2.2-3707.019 (West 2010); VA. CODE ANN. § 2.2-3711(A) (West 2010).

Washington: WASH. REV. CODE ANN. § 42.30.110(1) (West 2010).

West Virginia: W. VA. CODE ANN. § 6-9A-2(4) (LexisNexis 2009); W. VA. CODE ANN. § 6-9A-4(b) (LexisNexis 2009).

Wisconsin: WIS. STAT. ANN. § 19.85(1) (West 2010).

Wyoming: WYO. STAT. ANN. § 16-4-405(a) (2010).

THE MODERN HISTORY OF PROBABLE CAUSE

WESLEY MACNEIL OLIVER*

We often assume that those who wrote the Constitution understood its terms in a way that bears at least some similarity to the way we understand those terms today. This assumption is essential to the legitimacy of using Framing Era sources to inform the meaning of Constitutional provisions that regulate this system. This assumption is incorrect for one of the most important terms in criminal procedure. Probable cause meant something very different to the Framers than it means to modern lawyers. Probable cause was, as a practical matter, often nothing more than a pleading requirement for victims or officers who witnessed crimes. The modern notion of probable cause, an evidentiary threshold permitting a search or arrest that can be satisfied by the fruits of an officer's investigation, is a creation of the mid-nineteenth century. As with a number of Constitutional doctrines regulating the criminal justice system, we must look beyond the Framing Era to discover the origins of probable cause as it is understood by present-day lawyers.

INTRODUCTION 377
I. VICTIMS DROVE EIGHTEENTH-CENTURY CRIMINAL INVESTIGATIONS 381
A. Criminal Investigations 382
B. Customs Investigations 394
II. NINETEENTH-CENTURY PROHIBITION FASHIONED POLICE-FRIENDLY SEARCH STANDARD 399
III. LAW ENFORCEMENT NEEDS PROMPTED A BROADER ARREST POWER 419
CONCLUSION 428

INTRODUCTION

Our reliance on Framing Era materials to define how constitutional provisions ought to apply to modern law enforcement practices depends on

* Associate Professor, Widener University School of Law. B.A., J.D., University of Virginia; LL.M., J.S.D., Yale. This article is part of my dissertation on the nineteenth- and early twentieth-century development of modern criminal procedure. My thanks to the John H. Steiker Fund at Yale Law School for supporting portions of the research relating to Maine's nineteenth-century prohibitory law.

a very basic premise: that the Framers understood the terms of their document as we understand them today. For one of the most important constitutional terms regulating modern criminal procedure—probable cause—that simply is not the case.

In the Framers' victim-driven criminal justice system, probable cause was both more and less restrictive than it is under modern law. Probable cause was not enough to initiate a search or perform an arrest. Unless an officer saw a crime in progress, probable cause was sufficient for an arrest only if a victim attested that a crime had occurred. Officers were, therefore, most unlikely to act on mere suspicion, regardless of how strong it may be, lest they face civil damages. The Framing-Era criminal justice system did not, however, need to depend on officers investigating crimes and vigorously acting on their suspicions. An oath that a crime had occurred was all the evidence required for a victim to obtain a warrant to search for physical evidence in criminal cases. The victim merely had to assert that he had probable cause to suspect the person identified as the culprit, or had probable cause to believe evidence of a crime could be located in the identified location. Probable cause was essentially a pleading requirement that was easy for victims to satisfy but nearly impossible for public investigators in criminal cases to satisfy.

The reasons those in the Framing Era relied on victims were twofold. First, the eighteenth-century criminal justice system had little choice but to rely on victims.[1] The apparatus of law enforcement was in its infancy, ill-equipped to investigate criminal activity as a matter of routine.[2] Furthermore, its officers did not enjoy a privileged status in the social hierarchy. Therefore, eighteenth-century constables and watchmen lacked the capability and even the public trust necessary to engage in criminal investigations. Second, the eighteenth-century criminal justice system *could* rely almost exclusively on victims. Victimless crimes were virtually unknown in the Framing Era, so there was little need for the eighteenth-

---

1. *See* George C. Thomas, III, *Time Travel, Hovercrafts, and the Framers: James Madison Sees the Future and Rewrites the Fourth Amendment*, 80 NOTRE DAME L. REV. 1451, 1468–69 (2005); J.M. Beattie, *Early Detection: The Bow-Street Runners in Late Eighteenth-Century London*, *in* POLICE DETECTIVES IN HISTORY 1750–1950 15 (Clive Emsley & Haia Shpayer-Makov, eds., 2006) (observing that "there were severe limits as to the help victims of crime could expect to receive from [constables]."); 3 I.N.P. STOKES, THE ICONOGRAPHY OF MANHATTAN ISLAND 642–44 (1918).

2. Burt Neuborne, *The House was Quiet and the World was Calm The Reader Became the Book*, 57 VAND. L. REV. 2007, 2032 n.78 (2008) ("The civilian law enforcement authorities contemplated by the Founders did not include large professional police forces, which did not evolve until the middle of the nineteenth century. Instead, civilian law enforcement was the province of bounty hunters, individual officials, and/or *ad hoc* bodies, often using temporary personnel provided by powerful private interests or drawn from the local population."); *see generally* Wesley MacNeil Oliver, *Magistrates' Examination, Police Interrogations, and* Miranda*-Like Rules in the Nineteenth Century*, 81 TUL. L. REV. 777 (2007) (describing the rise of police interrogation).

century probable cause standard to authorize the intervention of the criminal justice system without a victim's complaint. Moreover, in a world in which victims were strictly liable for the fruitless searches they requested, there was little reason to require victims to provide more than their assurance that they had suspicion.[3]

The modern notion of probable cause, an evidentiary threshold that can be satisfied by anyone with relevant information, developed as society called for, and came to accept, modern police forces and began to regulate private moral practices. In no small part, metropolitan police forces were created for the express purpose of investigating and controlling crime.[4] The existence of these departments created pressure for a legal standard that did not require them to first ensure that a crime had been committed before arresting. The new standard developed despite the concerns created by the abuses of early police forces.

These new law enforcement organizations would have soon discovered that they also had an interest in a legal rule that would allow them to conduct *searches* without a victim's complaint. Mid-nineteenth century moral crusaders, however, beat them to the punch. Statewide versions of Prohibition preceded National Prohibition by about seventy years and required a search and seizure mechanism for enforcing this victimless crime. Unwilling to grant temperance zealots crime victims' power to

---

3. *See* Fabio Arcila, *The Death of Suspicion*, 51 WM. & MARY L. REV. 1275, 1318 (2010) (describing strict liability for affiant who sought fruitless search in the eighteenth century).

4. There is substantial agreement that the immediate impetus for these new departments was a wave of nineteenth-century riots, though the mission statements of these new departments all included the investigation of crime. *See* Robert Libman & Michael Polin, *Perspectives on Policing in Nineteenth Century America*, 2 SOC. SCI. HIST. 346 (1978) (reviewing scholarship on the creation of early police forces); ALLEN STEINBERG, THE TRANSFORMATION OF CRIMINAL JUSTICE: PHILADELPHIA, 1800–1880, at 119–20 (1989) ("Direct attempts at reform and efforts to retrain private prosecution made little contribution to the development of state prosecution. Instead, it emerged piecemeal, as a response to the increasing erosion of public order, primarily through the haphazard growth of the authority of the police."); SAMUEL WALKER, A CRITICAL HISTORY OF POLICE REFORM: THE EMERGENCE OF PROFESSIONALISM 4 (1977) (contending that modern police forces were developed as a "consequence of an unprecedented wave of civil disorder that swept the nation between the 1830s and the 1870s."); *see also* MARILYNN JOHNSON, STREET JUSTICE: A HISTORY OF POLICE VIOLENCE IN NEW YORK CITY 17–18 (2003). *But see* ERIC MONKKONEN, POLICE IN URBAN AMERICA, 1860–1920, at 56 (1981) (contending that cities seized the opportunity to create a mechanism of social control but were not motivated by any particular events); EDWIN G. BURROWS & MIKE WALLACE, GOTHAM: A HISTORY OF NEW YORK CITY TO 1898, at 637–38 (1999) (attributing to a brutal unsolved murder willingness of New Yorkers to finally accept a new police force); AMY GILMAN SREBNICK, THE MYSTERIOUS DEATH OF MARY ROGERS: SEX AND CULTURE IN NINETEENTH-CENTURY NEW YORK 87 (1995) (noting a brutal unsolved murder as a source for New Yorkers' willingness to accept a new police force).

satisfy probable cause upon a mere plea, legislators offered a new method for obtaining a search warrant in liquor cases. Prohibition statutes required applicants for search warrants to describe the liquor sales they allegedly observed in order to obtain warrants to search a dwelling. This was the first time a warrant could be obtained in an ordinary criminal case by an investigator who, though he could not say with absolute certainty that a crime had been committed, could satisfy probable cause, understood as an evidentiary threshold.

Therefore, probable cause, as we understand it today, is not the Framing-Era standard referred to in the Fourth Amendment.[5] Probable cause as an evidentiary threshold effectively did not exist in criminal cases in the late eighteenth century. The origins of the modern standard lie neither with the Framers, nor in ancient doctrines that long preceded their work. Modern probable cause—a standard for criminal cases—was a by-product of the work of mid-nineteenth-century reformers.

This article traces the mid-nineteenth-century development of this criminal standard believed to be of considerably older origins. Part I looks at the standard in the Framing Era, observing that there were two parallel tracks of law enforcement during this period. The enforcement of ordinary criminal laws depended on victims' complaints while customs and revenue enforcement could, obviously, not await the complaint of victim. In the early years of the country, these parallel tracks remained quite separate. Customs officials, who were no more harmed by violations than any other members of society, were necessarily required to obtain warrants on the basis of their investigations. Ordinary constables and watchmen, who enforced the general criminal law, could—and were expected to—rely on victims' investigations.

These systems began to merge, as Part II describes, as Prohibitionists sought a mechanism to search for alcohol. There was considerable distrust of those zealots who would seek warrants, prompting a mechanism to ensure the accuracy and veracity of complaints in the cases of victimless crimes. Probable cause, the evidentiary threshold sufficient for a search that we know today, developed as victimless crimes made a new method of authorizing searches a necessity. Distrust of Prohibition investigators ensured that this new standard for victimless searches would not rely on the

---

5. Many commentators have observed the increase in police officers' search and seizure powers beginning in the mid-nineteenth century with the creation of professional police departments, but no one has previously attempted to explain how feared rules giving extraordinary discretion to officers came to be accepted. *See* Thomas Y. Davies, *Recovering the Original Fourth Amendment*, 98 MICH. L. REV. 547, 552 (1999); Carolyn B. Ramsey, *In the Sweat Box: A Historical Perspective on the Detention of Material Witnesses*, 6 OHIO ST. J. CRIM. L. 681, 689 (2009) (noting nineteenth-century "shift toward greater police powers over the suspect"); David Alan Sklansky, *Is the Exclusionary Rule Obsolete?*, 5 OHIO ST. J. CRIM. L. 567, 579 (2008) (noting "dramatic" and "all encompassing" changes that accompanied creation of nineteenth-century police forces).

good faith of complainants as the criminal law had previously done. While the nineteenth-century version of Prohibition did not last, it left a search standard permitting magistrates to authorize search warrants on probable cause, as understood in the twenty-first century, alone.

A standard that allowed a government intrusion once an investigator had sufficient evidence was obviously useful to the work of members of newly created metropolitan police departments. As Section III describes, this standard allowed them to arrest immediately when their investigations suggested the guilt of a suspect. Such a standard was nearly essential for the new forces to perform the role assigned to them of aggressively preventing and solving crime. The early years of at least one police department, the New York Metropolitan Police Department, reveal that early and frequent misconduct made the public understandably reluctant to trust these new officers. The interests of the new police department nevertheless prevailed and the new arrest standard was embraced.

Probable cause as we understand it today, a foundational criminal law standard believed to substantially pre-date the Constitution, was thus not a criminal law standard at all in the eighteenth century. Rather, this standard, which alone justifies a search or arrest in a criminal case, is a creature of the mid-nineteenth century. At least in criminal cases, it meant something very different to the Framers than it means to modern lawyers. If history is to be a guide, its usefulness begins no earlier than the point at which our understanding of these terms began to map onto modern practice. For probable cause, that point occurred as law enforcement and Temperance interests first converged.

I. VICTIMS DROVE EIGHTEENTH-CENTURY CRIMINAL INVESTIGATIONS

The rules governing ordinary criminal investigations recognized and marginalized the role of the eighteenth century's part-time law enforcement officers in the criminal justice system. They were the ministerial assistants—the muscle—for victims and magistrates who directed their searches and seizures.[6] A victim's oath that a crime had occurred, and that he suspected a particular person, was both necessary and sufficient to initiate a criminal prosecution, leaving only a minor role for the constable.[7] Customs officers, by contrast, could act on the basis of what they learned

---

6. *See* Beattie, *supra* note 1, at 15; Roger Lane, *Urban Police in Nineteenth Century America*, 15 CRIM. & JUST. 1, 5 (1992); H.B. Simpson, *The Office of Constable*, 10 ENG. HIST. REV. 625, 635–36 (1895).

7. *See* Fabio Arcila, Jr., *In the Trenches: Searches and the Misunderstood Common-Law History of Suspicion and Probable Cause*, 10 U. PA. J. CONST. L. 1, 17–45 (2007) (contending that magistrates in the Framing Era did not require applicants for warrants to provide facts supporting their suspicions); WILLIAM J. CUDDIHY, THE FOURTH AMENDMENT: ORIGINS AND ORIGINAL MEANING, 602–1791 754, 757 (2009). *But see*, Davies, *supra* note 5, at 623.

through their investigations. Unlike ordinary officers, they routinely sought warrants and acted without warrants. Ordinary officers could not act on any quantum of proof—probable cause or otherwise—with or without a warrant, unless a crime had actually occurred, typically requiring them to wait for victims' complaints. Probable cause alone therefore had no role in the ordinary criminal justice system. Until broader search and seizure powers were conferred on officers enforcing the criminal law in the mid-nineteenth century, there were two very different schemes of search and seizure law in this country—one for criminal investigations, the other for customs and revenue enforcement.

A. Criminal Investigations

Early nineteenth-century criminal procedure severely limited the discretion of the majority of officers by effectively making them the ministerial assistants of magistrates and, ultimately, crime victims. Crime victims at the turn of the nineteenth century exercised the greatest discretion of any of the actors in the ordinary criminal justice system.[8] For most crimes, they alone conducted the investigation, identified suspects, and determined whether their suspicions were adequate to initiate a criminal prosecution.[9] Once victims announced their suspicions, constables were given fairly precise directions about the persons or property to seize.[10] Even

---

8. *See* William F. McDonald, *Towards a Bicentennial Revolution in Criminal Justice: The Return of the Victim*, 13 AM. CRIM. L. REV. 649, 650–54 (1975).

9. There seems to have been some variation in state practices, as one would logically expect in a world lacking modern instantaneous communication capability. Sources from some states suggest that applications for search or arrest warrants during the Framing Era required complainants to provide a factual basis for their suspicions. *See* Thomas Y. Davies, *Can You Handle the Truth? The Framers Preserved Common Law Criminal and Arrest and Search Rules in "Due Process of Law" – "Fourth Amendment Reasonableness" is Only a Modern, Destructive, Judicial Myth*, 43 TEX. TECH. L. REV. 51, 90–91 (2010) (observing that the Virginia Constitution of 1776 and the North Carolina Constitution of 1777 required that criminal warrants be supported by "evidence of a fact committed," while the Massachusetts Constitution of 1780 required only that "the cause or foundation" for a warrant be "supported by oath or affirmation."). As discussed below, however, it seems likely that the practice even in Virginia and North Carolina did not involve complainants providing the factual basis for their suspicions. In New York, for instance, even after a statute made a magistrate's duty to decide whether the facts offered by the complainant justified the warrant, in practice these magistrates do not have appear to have done anything other than accept the affiant's assertion that he had probable cause.

10. *See* Beattie, *supra* note 1, at 15; Lane, *supra* note 6, at 5. The broad power of search incident to arrest would seem to undermine this claim and, as a matter of pure doctrine, surely it does. *See* discussion *infra* at note 43. The victim, however, accompanied the officer and directed his search. *See* 5 RICHARD BURN, THE JUSTICE OF THE PEACE, AND PARISH OFFICER 199–200 (1776). This is not to say that there was not broad discretion in the eighteenth century to search for stolen goods. The discretion, however, was as a practical

when an officer had a sound basis for suspecting guilt, there was no mechanism for the officer to seek a warrant. An applicant for a warrant had to swear that a crime had been committed, which an officer could not do in most cases.[11] Before taking any action, a Colonial or early American officer responsible for enforcing the criminal law waited for a complainant to obtain a warrant, which shielded the officer from civil liability for fruitless searches or erroneous arrests.[12] Once the complaint was made, the officer relied on the victim's suspicions—he had no reason to conduct his own investigation.[13]

Victims exercised extraordinary discretion in this system. A criminal action at the turn of the nineteenth century was generally commenced by securing a warrant for a suspect's arrest or a warrant to search for particular property.[14] It was remarkably easy for crime victims to obtain arrest and

---

matter, exercised by the victim of the crime, not the officer.

11. *See* Davies, *supra* note 5, at 622–23.

12. The public had an intense fascination with search and seizure law at two points in American history: the era immediately preceding the American Revolution and during the effort to enforce national Prohibition. One of the critics of Prohibition, United States Senator A. Owsley Stanley of Kentucky (one of the country's largest producers of alcohol, then and now) observed that "the right to search and seize without a warrant was never vested in constables." A. Owsley Stanley, *Search and Seizure: Senator Stanley Attacks Constitutionality of New Prohibition Act*, N.Y. TIMES, Jan. 8, 1922, at 88. His conclusion was certainly correct with regard to the specific law enforcement officers to which he referred. *See* Davies, *supra* note 5, at 640–41. Customs and revenue officers, since the earliest days of the republic, however, had been vested with substantial powers of warrantless search and seizure. *See* Tracey Maclin, *Let Sleeping Dogs Lie: Why the Supreme Court Should Leave Fourth Amendment History Unabridged*, 82 B.U. L. REV. 895, 924 (2002) (observing that the modern Court has used the broad powers of customs agents to search without warrants to justify searches to enforce ordinary domestic crimes).

13. *See* discussion at *supra* note 12 and accompanying text.

14. Arrest warrants were far more common in the late eighteenth and early nineteenth centuries than search warrants. Search warrants were generally useful only in cases involving stolen items. In these collections, warrants in theft cases are more often for an arrest than for a search. ELIJAH ADLOW, THRESHOLD OF JUSTICE: A JUDGE'S LIFE STORY (1973) (describing Judge Adlow's discovery of these documents); Barrett Warrants (1787–1791), Gorham Warrants (1816–1818), Adlow Collection, Boston Public Library. Probable cause necessary to obtain a search warrant also permitted the applicant to obtain an arrest warrant. The very broad doctrine of search incident to arrest permitted an officer to search the arrestee's entire house for the stolen item. TELFORD TAYLOR, TWO STUDIES IN CONSTITUTIONAL INTERPRETATION 27–29 (1969); Morgan Cloud, *Searching Through History, Searching for History*, 63 U. CHI. L. REV. 1707, 1729 n.73 (1996) (describing Taylor's conclusion about a broad search incident to arrest doctrine as "noncontroversial").

Despite Cloud's conclusion, there has been some debate about the scope of the doctrine of search incident to arrest. William Cuddihy colorfully described the scope as follows. "Anyone arrested [in the eighteenth century] could expect that not only his surface clothing but his body, luggage, and saddlebags would be searched and, perhaps, his shoes, socks, and mouth as well." Thomas, *supra* note 1, at 1474 (citing CUDDIHY, *supra* note 7, at

search warrants, making public investigations unnecessary, at least in those cases in which the victim was fairly comfortable identifying the culprit. A complainant would appear before a magistrate and swear that a crime had occurred and that he had probable cause to believe the identified suspect guilty, or that evidence of the crime could be located in a particular location.[15] His assertion of the injury associated with the crime—i.e., loss of property in a theft case—was sufficient to demonstrate that the crime had occurred.

What it meant for the complainant to provide the magistrate probable cause has become the subject of a fairly intense debate in the academic community. Thomas Davies has argued that a complainant in the Framing Era was required to provide a magistrate with the facts upon which he based his suspicions and that the magistrate was to review the facts to determine whether they rose to the level of probable cause.[16] By contrast, Fabio Arcila has contended that, as a practical matter, probable cause was analogous to a pleading requirement.[17] He concludes that magistrates were not performing a gatekeeper function at all when presented with requests for warrants.[18] According to Arcila, a victim was only required to swear that a crime had been committed and that he *had* probable cause to believe the named suspect was guilty or that evidence of the crime could be discovered in the identified location.[19] Davies relies upon seventeenth- and eighteenth-century treatises that describe a magistrate as having a duty to consider the facts upon which the complainant relies for his suspicion. Arcila relies upon justice of the peace manuals and form books of the same period which appear to require the magistrate to ensure only that the complainant has sworn that a crime has occurred and that he, in fact, has

---

847–48). George Thomas has quite reasonably responded that a society that strictly limited an officer's right to arrest a suspect would be reluctant to allow an officer to search "beyond what was necessary to disarm him." *Id.* While it is hard to argue with Thomas' logic, his conclusion seems undermined by the few archived warrants that have survived from the late eighteenth and early nineteenth centuries in the Adlow Collection of the Boston Public Library. Victims of theft would have been interested in securing the evidence necessary to prove that the theft occurred and, far more importantly, ensuring the return of their property. If the search incident to arrest power were not quite so broad, one would expect search warrants rather than arrest warrants to have been issued in the vast majority of theft cases.

15. Form of a Complaint to Obtain a Search Warrant, *in* GENTLEMEN OF THE BAR OF NEW YORK, THE ATTORNEY'S COMPANION 435 (Poughkeepsie, N.Y., P. Potter & S. Potter 1818).

16. Davies, *supra* note 5, at 651–52; Thomas Y. Davies, *An Account of* Mapp v. Ohio *that Misses the Larger Exclusionary Rule Story*, 4 OHIO ST. J. CRIM. L. 619, 621–22, 624 n. 19 (2007).

17. *See* CUDDIHY, *supra* note 7, at 582 ("The general rule was that magistrates neither examined complainants independently to determine their adequacy for warrants nor withheld warrants if the assessment was negative."). *See generally* Arcila, *supra* note 7.

18. *See generally* Arcila, *supra* note 7.

19. *See generally id.*

probable cause. And while Davies points out that the form books contained explanatory notes reiterating the duty of the magistrate to determine that probable cause existed,[20] it is the forms themselves, not the explanatory notes that followed, that appear to have driven the practice. As one might expect, government officials appear to have developed a practice of obtaining merely the information required to complete the forms.

Legal treatises dating back to the seventeenth century observed that magistrates were to examine the facts supporting an application for an arrest or search warrant. This rule was announced by such legal luminaries as Matthew Hale, William Hawkins, and William Blackstone.[21] There are, nevertheless, substantial reasons to believe that Arcila has accounted for the actual practice of eighteenth- and early nineteenth-century magistrates.

In their landmark work on law enforcement in colonial New York, Julius Goebel and T. Raymond Naughton observe that magistrates around the turn of the eighteenth century occasionally declined to issue warrants requested of them, but that there were frequent complaints made that magistrates felt they had such discretion.[22] A magistrate's review of the facts supporting a complainant's suspicion seems to have been an aberration and there was great public pressure to eliminate these aberrations. Further, if magistrates at the turn of the nineteenth century were requiring complainants to provide factual support for their suspicions, they made no record of these facts. The few actual warrant applications that have survived from the turn of the nineteenth century reveal that, consistent with the form books Arcila cites, warrant applications contained no recitation of the facts complainants relied upon.[23]

Several pieces of evidence from the mid-nineteenth century provide further support for the conclusion. Oliver Barbour's treatise on New York criminal procedure observed in 1841 that, "[a]t common law, it seems a magistrate might issue his warrant upon a general oath of suspicion merely. This was on the ground that the complainant was a competent judge of the matters upon which his suspicion rested."[24] Henry Dutton's Connecticut

---

20. *See* Davies, *supra* note 9 at 78 n.122.

21. *See* 4 WILLIAM BLACKSTONE, COMMENTARIES ON THE LAWS OF ENGLAND 290–91 (London, 1826) (1790); 2 MATTHEW HALE, THE HISTORY OF THE PLEAS OF THE CROWN 150 (London, E. Rider, Little Britain 1800); 2 WILLIAM HAWKINS, A TREATISE ON THE PLEAS OF THE CROWN 130–31 (Thomas Leach ed. 6th ed., London, His Majesty's Printer 1824) (1787).

22. JULIUS GOEBEL, JR. & T. RAYMOND NAUGHTON, LAW ENFORCEMENT IN COLONIAL NEW YORK 424–25 (1944).

23. *See* Barrett Warrants (1787–1791), Gorham Warrants (1816–1818), *supra* note 14; Thomas Y. Davies, *Revisiting the Fictional Originalism in* Crawford*'s "Cross-Examination Rule"- A Reply to Mr. Kry*, 72 BROOK. L. REV. 557 (2007).

24. OLIVER L. BARBOUR, THE MAGISTRATE'S CRIMINAL LAW: A PRACTICAL TREATISE ON THE JURISDICTION, DUTY AND AUTHORITY OF THE JUSTICES OF THE PEACE IN THE STATE OF NEW-YORK, IN CRIMINAL CASES 454 (1841). Barbour's conclusion that New York state practice after the adoption of the Revised Laws of 1829 required an applicant to demonstrate

treatise observed that a "justice of the peace may issue a warrant to search for stolen goods; but to authorize this, there must be the oath of the applicant that the goods have been stolen, and that he strongly suspects that they are concealed in a certain place."[25] As a Justice of the Connecticut Supreme Court, Dutton would observe that the "oath of a person who lost the goods" swearing that he "has just grounds to suspect and does suspect that the goods were taken by [the identified culprit]" was sufficient to obtain a search warrant.[26]

Finally, mid-nineteenth-century courts were obsessed about the specific language in the complaint, which was generally contained in the pre-printed portion of a form that the complainant filled out.[27] State courts found search and arrest warrants invalid because the complainant had sworn that he "had cause to suspect and did suspect" that the identified person was the culprit, or that evidence of his crime could be found in a particular location.[28] The

---

the factual basis of his suspicions to a magistrate was demonstrated to be false—at least as a practical matter—with the reports of the Commissioner on Pleading and Practice, who observed that magistrates were not examining the factual foundation at all. *See* discussion *infra* notes 36–37 and accompanying text.

25. HENRY DUTTON, A REVISION OF SWIFT'S DIGESTS ON THE LAWS OF CONNECTICUT 505 (New Haven, Durrie & Peck 1851). For a more complete description of Henry Dutton's life, see discussion *infra* note 26.

26. Lowrey v. Gridley, 30 Conn. 450, 456–57 (Conn. 1862). Connecticut is admittedly not a typical case. Henry Dutton had a clear motive to resolve any ambiguity about a magistrate's duty to examine the facts supporting a complainant's requested warrant in favor of not requiring such an examination. Dutton, a Yale Law Professor in 1851, wrote his treatise prior to advocating passage of the state's prohibitory law as a member of the legislature. The law he advocated contained the most permissive search standard in the country. *See* YALE UNIVERSITY, OBITUARY RECORD OF GRADUATES OF YALE COLLEGE: DECEASED FROM JULY, 1859 TO JULY, 1870 (New Haven, Tuttle Morehouse & Taylor 1870); Letter to the Editor, HARTFORD COURANT, April 21, 1854, at 2 (describing Dutton's role); *The Maine Liquor Law – As just passed by the Connecticut Legislature*, NEW YORK TIMES, June 22, 1854 (describing law). The mere allegation of three persons that liquor was present in a home was sufficient to obtain a warrant under this statute. *Id.* The same legislature that enacted the prohibitory law, elected him Governor of Connecticut. He subsequently became a justice on the state supreme court, where in the *Lowrey* case he was asked to pass on the constitutionality of the search and seizure process permitted under the statute. Dutton reasoned that the liquor law was more protective of individual liberty than searches for stolen goods, which could proceed on the mere allegation of a single person that the goods were in a particular location. *Lowrey*, 30 Conn. at 456–57. One could reasonably surmise that Dutton was laying the legal groundwork for the *Lowrey* decision for over a decade, even though it would be a stretch to suppose that he saw himself writing it ten years later.

27. Form of a Complaint, *supra* note 15; 2 N.Y. REV. STAT., pt. 4, ch. 2, tit. 2 (1836); Mass. Rev. Stat. title II, ch. 142, § 1 (1836).

28. *See* Thomas Y. Davies, *The Fictional Character of Law and Order Originalism: A Case Study of the Distortions and Evasions of Framing Era Arrest Doctrine in* Atwater v. Lago Vista, 37 WAKE FOREST L. REV. 239, 381 n.480 (2002) (citing Humes v. Taber, 1 R.I. 464, 465 (1850)). The Massachusetts Supreme Judicial Court similarly found that a search

courts found these warrant applications insufficient because the complainant had not demonstrated adequate certainty about his suspicions.[29] Complainants in these states were thus required to swear that they had "probable cause to believe and did believe" that the identified person was the culprit, or that evidence of his crime could be located in the location identified.[30] If magistrates in the mid-nineteenth century reviewed the facts complainants offered, the complainant's characterization of his level of suspicion would have been irrelevant. The magistrate's independent determination that there was probable cause would have overcome any lack of certainty expressed in the form pleading used by the complainant.

Magistrates certainly could—and did—reject warrant applications, but their rejections appear to have been based on concerns about complainants, not their complaints. This victim-driven system was willing to trust victims only so long as they appeared trustworthy. As one treatise writer observed, "Where a magistrate has reasonable ground to believe that the charge preferred is the offspring of malice and a corrupt heart, he may require further evidence of its truth than the oath of the complainant."[31] Somewhat

---

warrant for lottery tickets based on a complainant's oath that he had "probable cause to suspect" the tickets present was insufficient. Commonwealth v. Certain Lottery Tickets, 59 Mass. 369, 372 (Mass. 1850). Like Rhode Island, Massachusetts, by statute, required an applicant for a search warrant to swear that he believed the evidence could be discovered in the location identified. The court held that a warrant application "sworn to in the old form," i.e., the one used before the Massachusetts Revised Statutes of 1837, was invalid, at least for searches that had not previously been authorized on "suspicion" rather than "belief." *Id.* at 372. The Massachusetts case offered something of a preview of issues that would arise with Prohibition. Massachusetts, unlike most states of the mid-nineteenth century, had passed a statute authorizing a search for evidence of victimless crimes. Searches could be instituted in Massachusetts for counterfeit money, obscene publications, lottery tickets, or gaming devices. Mass. Rev. Stat. tit. II, ch. 142 § 2. The Revised Statutes required a complainant to assert his "belief" in seeking all warrants, whether to search for stolen goods or evidence of the new victimless crimes. *Id.* at § 1. The court held that it was not required to consider, in this case involving lottery tickets, whether the old form was adequate for a search warrant to recover stolen goods. The suggestion that different standards might apply to searches for victimless crimes would reappear when Prohibition created a realistic threat that victimless crimes would be prosecuted.

29. *See Certain Lottery Tickets*, 59 Mass. at 372.

30. *Id.*

31. *See* JOHN C.B. DAVIS, THE MASSACHUSETTS JUSTICE: A TREATISE UPON THE POWERS AND DUTIES OF JUSTICES OF THE PEACE: WITH COPIOUS FORMS 192 (Worchester, Mass. 1847) (also observing that the magistrate "may, also, upon deliberate consideration, refuse to institute a criminal process," suggesting that usual course was to grant requested warrant); *see also* BARBOUR, *supra* note 24, at 451 (stating that a magistrate "ought not . . . to proceed upon a complaint solely because such complaint has been made; for though there be a positive charge on oath by a competent witness, if the justice sees that no credit is to be given to it, he may, and should doubtless, decline acting on it"). Barbour's treatise did recognize that New York judges had a duty to inquire into the facts supporting warrant applications. BARBOUR, *supra* note 24, at 454. This reference was, of course, to the

remarkably, victims were never required to provide any sort of surety when they requested arrest or search warrants. This is particularly striking in light of the fact that in the eighteenth century some authorities concluded that applicants for warrants were strictly liable in trespass for erroneous arrests or fruitless searches.[32] By the mid-nineteenth century, the burden had shifted. A victim of an improper search or arrest had the burden of proving that the complainant lacked probable cause, virtually immunizing him from suit.[33]

The mid-century battle over state prohibitory laws provides further evidence that magistrates were not expected to scrutinize warrant applications. As will be discussed much more fully below, legislatures refused to authorize warrants to search for liquor that followed the same procedures used for ordinary search and arrest warrants.[34] Prohibition bills permitting these warrants were accepted only after they were modified to require their applicants to explain why they believed alcohol could be discovered in the location indicated and a magistrate to find these facts provided probable cause.[35] If this procedure merely restated existing practice, it seemingly could not have ameliorated the concerns of even a single opponent of the proposed law.

Whatever the actual practice at the turn of the nineteenth century, it is very clear that by the middle of the century, magistrates were not considering the grounds supporting a requested warrant, even when expressly required to do so by statute. Statutory revisions in New York in 1829, and Massachusetts in 1836, contained provisions requiring a judge's evaluation of the facts supporting a complainant's fact, but magistrates ignored both provisions.[36] The New York Commissioners on Pleading and

---

requirement codified in New York's statutes of 1829, which would not be followed in practice. As a corollary, the reputation of the suspect was expressly identified as a sufficient basis for a magistrate to determine that the complainant had demonstrated probable cause. *See* Joseph D. Grano, *Probable Cause and Common Sense: A Reply to the Critics of* Illinois v. Gates, 17 U. MICH. J.L. REFORM 465, 481 n.94 (1983–84) (quoting William Hawkins' eighteenth-century treatise).

32. *See* Entick v. Carrington, (1765) 95 Eng. Rep. 807 (K.B.); 2 Wilson 275 (Eng.).

33. *See* Burns v. Erben, 40 N.Y. 463, 465 (N.Y. 1869) (in an action for malicious prosecution, "the burden was upon the plaintiff to show a want of probable cause.").

34. *See* discussion *infra* notes 114–133 and accompanying text.

35. *See* discussion *infra* notes 115–133 and accompanying text.

36. This is one example of the notice taken of New York's criminal procedure outside the Empire State. The Massachusetts legislature used virtually the identical language that the New York Legislature had used. GEORGE EDWARDS, A TREATISE ON THE POWERS AND DUTIES OF JUSTICES OF THE PEACE AND TOWN OFFICERS IN THE STATE OF NEW YORK, UNDER THE REVISED STATUTES WITH PRACTICAL FORMS (Ithaca, Mark, Andrus & Woodruff, 3rd ed. 1836) (quoting N.Y. Rev. Stat. p. 746, Tit. 7, Part IV, § 25 ("if such magistrate be satisfied that there is reasonable ground for [the complainant's] suspicion, he shall issue a warrant to search for such property.")); Mass. Rev. Stat. tit. II, ch. 142, §1 (1836) ("if [the magistrate] be satisfied that there is reasonable cause for [the complainant's] belief, [he] shall issue a

Practice, charged with the duty of producing a Code of Criminal Procedure, complained in 1850 of the "loose practice" of magistrates in issuing warrants. They observed that "[i]t is very common, for example, to state in cases of larceny, nothing more, than that the property was stolen taken away &c., by the person charged."[37]

There were no penalties for ignoring statutory provisions requiring magistrates to assess the strength of the facts supporting a victim's allegation of probable cause. As an example of this, the Massachusetts Supreme Judicial Court held in 1841 that there was no remedy against a magistrate for granting a warrant without considering the facts supporting the complainant's allegation, or against the complainant for requesting it[38]:

> The great security of the citizen from unreasonable arrest or seizure of goods is this, that the warrant is only to issue upon the oath of the complainant alleging a larceny, &c., and his belief that the party accused is guilty of the offence; or, in the case of seizure on a search warrant, that he believes the property stolen, embezzled, &c., to be in the place searched.[39]

This system placed great trust in victims who, by the mid-nineteenth century, were liable for malicious prosecution only if the target of the investigation could demonstrate that the complainant lacked probable cause.[40] When a crime victim went to the magistrate, he sought one of two types of warrants to initiate a criminal action. If he swore he knew who had committed the crime, he requested a warrant for the culprit's arrest; if he swore he knew where stolen goods could be found, he asked for a search warrant. For a variety of reasons, arrest warrants were far more common than search warrants.[41] In a world before forensic science, the only type of search warrant that would have been useful to a victim was one to recover stolen goods.[42] Further, the doctrine of search incident to arrest was extremely broad in the late eighteenth and early nineteenth centuries. If a victim swore he had probable cause to believe a particular person had stolen

---

warrant to search for such property.").

37. SELECT COMMITTEE ON CODE OF CRIMINAL PROCEDURE, REPORT, DOCUMENTS OF THE ASSEMBLY OF THE STATE OF NEW YORK, 78th Sess., NO. 150, at 79, § 149 (1855). [hereinafter CRIMINAL CODE].

38. Stone v. Dana, 46 Mass. (5 Metc.) 98 (1842).

39. *Id.* at 109–110.

40. *See* discussion at *supra* note 33.

41. *See e.g.*, Barrett Warrants (1787–1791), Gorham Warrants (1816–1818), *supra* note 14.

42. It is frequently stated that the common law only permitted searches for stolen goods, which is true, but there is a caveat: certain statutes allowed searches for smuggled goods and dangerous items, such as gunpowder or diseased or infected animals. Akhil Reed Amar, *Fourth Amendment First Principles*, 107 HARV. L. REV. 757, 765 (1994).

his property, an arrest warrant would have permitted his apprehension as well as a search of his premises for the missing items.[43]

State officers exercised almost no discretion in the investigatory or prosecutorial process. The constable's role in the criminal case ended with the arrest and any search that accompanied it. The magistrate was the only participant in the criminal justice system expected to question suspects.[44] The constable was not expected to question the suspect.[45] Some English authorities at the turn of the nineteenth century actually forbid the constable to question the suspect, though American authorities never adopted this position.[46] In the United States, custom and lack of institutional incentive were likely sufficient to prevent the practice of routine police interrogation from developing before the creation of professional police departments.[47] Beyond this, any inducement or promise held out by a constable threatened the admissibility of the statement the accused made further discouraging any effort at interrogation.[48]

Limitations on the officers power to investigate criminal matters, either pre-arrest or post-arrest were as much a function of customary practice, institutional incentives (or a lack thereof), and the constable's social standing (or lack thereof).[49] The history of a particularly despised type form of investigatory authority illustrates this point. Though they no longer existed at the time the United States Constitution was written, general

---

43. *See* TAYLOR, *supra* note 14, at 27–29.

44. *See* Albert W. Alschuler, *A Peculiar Privilege in Historical Perspective: The Right to Remain Silent*, 94 MICH. L. REV. 2625 (1996) (observing inconsistency between Fifth Amendment right to remain silent and routine practice of magistrates to interrogate suspects).

45. *See* Steven Penney, *Theories of Confession Admissibility: A Historical View*, 25 AM. J. CRIM. L. 309, 323 (1998) (observing that the duties of sheriffs, constables, and watchmen "did not generally include either the investigation of unsolved crimes or the interrogation of suspects").

46. *See* 1 S. MARCH PHILLIPPS, TREATISE ON THE LAW OF EVIDENCE 406 (New York, Banks, Gould & Co., 3rd ed.1849) (citing Rex v. Wilson (1817) 171 Eng. Rep. 353, 353; 7 Holt 596.). ("A confession, obtained without threat or promise has been received, notwithstanding it was elicited by a police officer.") Phillips and Amos noted, however, that there were English authorities to the contrary of this proposition. *Id.* (citing Rex v. Wilson (1817) 171 Eng. Rep. 353, 353; 7 Holt 596).

47. There were rare instances of officers performing interrogations and likely involved cases involving rewards. *See* Oliver, *supra* note 2, at 795 n.100, 797 (2007).

48. By the mid-nineteenth century, there was a growing consensus in England that any police interrogation rendered a statement involuntary. *See* Bram v. United States, 168 U.S. 532, 556 (1897) (describing English voluntariness rule).

49. Justice White concluded that the primary limitation on eighteenth-century officers was "the generally ministerial nature of the constable's office at common law." Payton v. New York, 445 U.S. 573, 607 (1980) (White, J., dissenting). Justice White was incorrect in concluding that "the constable possessed broad inherent powers to arrest," but his assessment of the institutional limits on early officers is certainly correct. *Id.*

warrants, permitting the officer to search wherever he suspected, or arrest anyone he suspected of committing the crime in question, had been fairly common in early colonial era.[50] Yet the American colonists were never outraged by the authority exercised by constables when they used these warrants.[51] The reason for this was that that in practice, constables bearing these general warrants searched those only places, and arrested only those persons identified by the complaining victims who sought warrants.[52] Controversies over general warrants in customs enforcement led to their quiet replacement with specific warrants in criminal cases. As will be discussed below, general warrants for customs searches were much less comfortably tolerated.[53] When ordinary officers were entrusted with great discretion, they were practically prevented from exercising it; the informal limits were as important as the formal.[54]

The formal limits were, however, far from insignificant. An officer's authority to act without a warrant prior to the mid-nineteenth century was quite limited.[55] He had authority to search without a warrant only incident to arrest, and only under very narrow circumstances could an officer arrest without a warrant.[56] If an officer actually witnessed a crime occur, he was

---

50. CUDDIHY, *supra* note 7, at 555 (outside Massachusetts and New Hampshire, "general warrants for stolen objects remained in use despite the opposition of legal authors.").

51. *Id.* at 575 ("Until the 1760s . . . colonial law had neither rejected general warrants nor embraced specific ones.").

52. *Id.* at 754, 757. A warrant issued after Independence illustrates the practice well. The constable was instructed: "You are commanded forthwith to search all suspected places and persons that the complainant thinks proper, to find his lost pork, and to cause the same, and the person with whom it shall be found, or suspected to have taken the same, and have him appear before some proper authority, to be examined according to law." Frisbee v. Butler, 1 Kirby 213, 213–14 (Conn. 1787).

53. *See* discussion *infra* notes 64–74 and accompanying text.

54. *See* Beattie, *supra* note 1 (observing that "there were severe limits as to the help victims of crime could expect to receive from [constables]."); Thomas Y. Davies, *Farther and Farther from the Original Fifth Amendment: The Recharacterization of the Right Against Self-Incrimination as a "Trial Right" in* Chavez v. Martinez, 70 TENN. L. REV. 987, 1004 (2003) ("The constable had neither a duty nor the authority to investigate the possibility of uncharged crimes; in fact, in the absence of a warrant, the constable had little more arrest authority than any other person."); JAMES F. RICHARDSON, THE NEW YORK POLICE: COLONIAL TIMES TO 1901 17–18 (1970) (describing similar powers by New York constables in early nineteenth century); H.B. Simpson, *The Office of Constable*, 10 ENG. HIST. REV. 625, 635–36 (1895) (distinguishing the role of modern police from constables who were "regarded merely as . . . police officer[s] attendant on the justices [of the peace] and other ministers of the crown."). For a description of the weakness of constables before the colonial era, see Joan Kent, *The English Village Constable, 1580–1642: The Nature and Dilemmas of the Office*, 20 J. BRIT. STUD. 26 (1981).

55. *See* Davies, *supra* note 5, at 554 ("At common law, controlling the warrant did control the officer for all practical purposes.").

56. Customs officers were permitted to search ships without warrants, but this

not liable if the person he arrested was not the culprit. Finally, an officer was immune from liability if he arrested an innocent person but a crime had in fact occurred and the officer had probable cause to believe the person he arrested had committed it. These bases for arrest also necessarily depended on a victim in most cases. Unless the officer witnessed the crime occur, he would seldom have a sufficient basis for concluding—at the risk of civil judgment—that a crime had in fact occurred. The part-time, often volunteer officers of the Framing Era, with little incentive to patrol, would have seldom discovered a crime in progress.[57]

Greater arrest powers in the early 1800s served only to bolster the importance of victims. Beginning in the early nineteenth century, an officer was not liable for false arrest if a citizen complained of a crime and identified the suspect for the officer to arrest.[58]

The lack of trust eighteenth-century society was willing to place in officers—and the lack of financial investment it was willing to make in police organizations—often left those least able to vindicate their rights solely responsible for doing so. By contrast, victims at the turn of the nineteenth century, who did not always know who had perpetrated the crimes against them, could have been aided by constables in the investigation. There was, however, no incentive for the constables to assist, because investigation was not regarded as part of their job. Wealthy victims could offer rewards, which obviously changed the incentives for officers, essentially rendering them private investigators.[59] In all other cases, the

---

warrantless search authority did not extend to constables seeking evidence of crimes. *Id.* at 571. At common law, officers were immune from liability for false arrest if they witnessed the crime occur, responded to a "hue and cry," had a warrant for the arrest, or had probable cause to believe the arrestee had committed a felony when a crime had in fact occurred. *See* Horace L. Wilgus, *Arrest Without a Warrant*, 22 MICH. L. REV. 798, 809 (1923–24). Officers were also permitted to arrest suspects for misdemeanors that were committed in their presence. *Id.* at 814.

57. *See* Wesley MacNeil Oliver, *The Neglected History of Criminal Procedure, 1850–1940*, 62 RUTGERS L. REV. 447, 451–452 (2010) (describing institutional incentives of Framing-Era officers).

58. *See* Davies, *supra* note 5, at 635–36. *See generally* Wilgus, *supra* note 56 (describing English and American arrest rules).

59. In the early years of the American republic, rewards were typically given by private parties seeking the return of their property *See* City Bank v. Bangs, 2 Edw. Ch. 95 (1833) (discussing circumstances under which officer was entitled to receive private reward); BURROWS & WALLACE, *supra* note 4, at 637; ROGER LANE, POLICING THE CITY: BOSTON, 1822–1885 56 (1967); RICHARDSON, *supra* note 54, at 62–63. In England beginning in the 1730s, the government offered rewards for the identification and successful prosecution of those committing more serious property crimes. *See* GERALD HOWSON, THIEF-TAKER GENERAL: THE RISE AND FALL OF JORDAN WILD (1970); JOHN H. LANGBEIN, THE ADVERSARY ORIGINS OF CRIMINAL TRIAL 109 (2003); Ruth Paley, *Thief-takers in London in the Age of the McDaniel Gang c. 1745–1754, in* DOUGLAS HAY & FRANCIS SNYDER (ED.), POLICING AND PROSECUTION IN BRITAIN 1750–1850 (1989).

public apparatus of the criminal justice system assisted the victim only after he made his complaint. The constable aided the victim in conducting any search and the magistrate interrogated the suspect, but any legwork, either prior to the allegation or trial, was left up to the victim.[60]

There were a couple of exceptions to this system of victim-driven investigations, neither of which involved conferring discretion on the constabulary. Customs and revenue violations, discussed below, could not have been investigated by victims because there were no victims.[61] Murders, for an entirely different reason, had to be investigated by someone other than the victim. Coroners, who were not required to have any medical training, assembled a jury that functioned much like modern grand juries.[62] These were private citizens who subpoenaed witnesses and considered any physical evidence, such as the crime scene and the body itself.[63] There were no public-employed criminal investigators until mid-nineteenth-century reforms created modern police officers, who would reluctantly be given the prerogatives of customs officers.

The Framing-Era and early American criminal justice systems did not involve investigations by public officers. Probable cause as it is understood in modern times would have allowed officers to act on the information they learned—or at least to seek judicial authorization to act on the information they learned. The eighteenth-century criminal justice system—driven by the investigations of private complainants—had no need for such a standard. A standard that allowed searches and seizure on the basis of an officer's investigation was, however, essential to customs and revenue enforcement in this era.

---

60. *See* United States v. Dickerson, 530 U.S. 428, 435 n.1 (2000) (observing that "custodial police interrogation is relatively recent because the routine practice of such interrogation is itself a relatively new development"); Alschuler, *supra* note 44.

61. *See* Thomas Y. Davies, *Correcting Search-and-Seizure History: Now-Forgotten Common-Law Warrantless Arrest Standards and the Original Understanding of "Due Process of Law"*, 77 MISS. L.J. 1, 118 (2007) (observing that John Adams, the primary drafter of the Massachusetts Constitution, doubtlessly recognized that the Virginia Constitution's requirements for a search warrant could not be satisfied for customs searches as no victim could swear that a crime had occurred).

62. *See* 1 JOSEPH CHITTY, A PRACTICAL TREATISE ON THE CRIMINAL LAW 158–59 (London, A.J. Valpy 1816) (observing unique method of investigation for murders); JULIUS GOEBEL, JR., CASES AND MATERIALS ON THE DEVELOPMENT OF LEGAL INSTITUTIONS 59 (1946) (describing process of murder investigation); JOHN IMPEY, THE OFFICE OF SHERIFF, SHEWING ITS HISTORY AND ANTIQUITY 440–41 (London, W. Clark and Sons 4th ed. 1817) (discussing duties of the coroner); George C. Thomas, III, *Colonial Criminal Law and Procedure: The Royal Colony of New Jersey 1749–1757*, 1 N.Y.U. J.L. & LIBERTY 671, 679–80 (2005) (describing a proceeding under coroner's inquest).

63. For a history of the coroner's jury, see Thomas A. Green, *The Jury and the English Law of Homicide, 1200–1600*, 74 MICH. L. REV. 413, 422–25 (1976); Irvin L. Langbein, *The Jury of Presentment and Coroner*, 33 COLUM. L. REV. 1329 (1933).

B. Customs Investigations

In contrast with ordinary law enforcement officers, customs officers had considerable discretion to initiate investigations and substantial financial incentives to do so. Even though the authority exercised by customs officers had been a major source of contention between Great Britain and the American colonies, customs and revenue officers continued to possess a unique type of discretion for decades after independence.[64] Customs officials were the only officers capable of seeking warrants—or engaging in warrantless searches or seizures—on the strength of facts they learned through their investigations.[65] Probable cause, as understood in the modern world, was a standard uniquely applicable to customs officers. The method of compensating customs officers, however, gave them considerably more incentive to search than modern police officers possess. They received a portion of the government's fee or forfeiture for a violation.[66]

As there was no victim to swear that a violation had occurred, the limits imposed on searches and seizures in ordinary criminal cases would have completely prevented the enforcement of customs and revenue laws if extended in those contexts. Eighteenth-century Americans' historical

---

64. *See* Davies, *supra* note 28, at 605–08 (observing that Framers did not intend Fourth Amendment to create the same search and seizure standard for customs officers and officers who enforced ordinary criminal laws).

65. The broad authority of eighteenth-century customs officers has been frequently recognized. *See* Carroll v. United States, 267 U.S. 132, 151 (1925) (describing authority of late eighteenth- and early nineteenth-century customs officers to search vessels without a warrant; Fabio Arcila, Jr., *The Framer's Search Power: The Misunderstood Statutory History of Suspicion & Probable Cause*, 50 B.C. L. REV. 363, 363, 410 (2009) (observing that early Congresses gave customs officers considerable immunity from suit); Richard H. Fallon, Jr. & Daniel J. Meltzer, *New Law, Non-Retroactivity, and Constitutional Remedies*, 104 HARV. L. REV. 1731, 1783 n.279 (1991) (describing Marshall Court cases providing customs officers immunity from suit for seizing goods without warrant even when there was no good faith basis for officers' suspicion of illegal conduct); Alfred S. Martin, *The King's Customs: Philadelphia, 1763–1774*, 5 WM. & MARY Q. 201 (1948) (describing the roles of customs officers in the Port of Philadelphia and the specific job performed by one officer whom the author regarded to be particularly honest and efficient); *see also* Amar, *supra* note 42, at 766 (describing authority of customs officers under late eighteenth-century statutes to search vessels without a warrant and suggesting that a warrant may not have been required under language of early customs laws to search homes, buildings, or stores). *But see* Tracey Maclin, *The Complexity of the Fourth Amendment*, 77 B.U. L. REV. 925, 952–53 (1997) (disputing Amar's claim). *See generally* THOMAS C. BARROW, TRADE AND EMPIRE: THE BRITISH CUSTOMS SERVICE IN COLONIAL AMERICA 1660–1775 (1967).

66. *See* Davies, *supra* note 5, at 659; *see also* Eric Blumenson & Eva Nilsen, *Policing for Profit: The Drug War's Hidden Economic Agenda*, 65 U. CHI. L. REV. 35 (1998) (describing modern issues with allowing police departments to retain a portion of forfeited funds).

concerns about the discretion of customs officers may, however, make the amount of discretion uniquely vested in these officers seem surprising.

General warrants empowering customs officers to search anywhere they suspected they would discover violations of import and tax laws particularly drew the ire of colonists. The first, and certainly most famous, of these controversies occurred in Massachusetts after the Superior Court issued customs officers a particular version of general warrants, known as writs of assistance, to discover evidence of illegal trading with French Canada.[67] When the warrants expired upon the death of King George II in 1760, Boston area merchants and smugglers (groups with largely overlapping memberships) retained two of the best lawyers in Massachusetts, Joseph R. Frese and James Otis, to argue against reissuing them.[68] Otis' now-famous argument objected that these warrants placed the liberty of every man in the "hands of a petty officer."[69] John Adams, then a young lawyer, was in the audience and later said of the argument against this sort of authority for customs officers, "Then and there was the child Independence born."[70]

The threat uniquely posed by customs officers brought about this intense criticism. The general warrant was not new to colonists in 1760. Constables had been issued general warrants throughout the colonial era to arrest unnamed persons or search unidentified places. While specific warrants, identifying the place to be searched or person to be seized, had come to replace general warrants in Massachusetts (though not in other colonies) well before Otis made this argument, this transition had been gradual and had not been provoked by any particular outrage.[71] By contrast with the Americans, the English had long had philosophical objections to general warrants. Treatise writers had regarded them to be illegal for over a

---

67. *See* John M. Burkoff, *"A Flame of Fire": The Fourth Amendment in Perilous Times*, 74 MISS. L.J. 631, 634–35 (2004).

68. Writs of Assistance conferred their extraordinary discretion on the officers to whom they were issued for the life of the sovereign in whose name they were issued. To be precise, they expired six months after, in this case, the death of King George II. *See* NELSON B. LASSON, THE HISTORY AND DEVELOPMENT OF THE FOURTH AMENDMENT TO THE UNITED STATES CONSTITUTION 57 (1970). The reputation of Boston's most prominent eighteenth-century merchants as smugglers has been well confirmed. *See* JOHN W. TYLER, SMUGGLERS AND PATRIOTS: BOSTON MERCHANTS AND THE ADVENT OF THE AMERICAN REVOLUTION (1986). *But see* O.M. Dickerson, *John Hancock: Notorious Smuggler or Near Victim of British Revenue Racketeers?*, 32 MISS. VALLEY HIST. REV. 517 (1946).

69. Otis' argument was recorded by a young John Adams who sat in the audience with a number of luminaries of the Boston legal, political, and commercial world. LASSON, *supra* note 68, at 58–59. Adams would later say of Otis' argument, "there the child of Independence was born." T. H. Breen, *Subjecthood and Citizenship: The Context of James Otis's Radical Critique of John Locke*, 71 NEW ENG. Q. 378, 378 (1998).

70. M. Blane Michael, *Reading the Fourth Amendment: Guidance from the Mischief that Gave It Birth*, 85 N.Y.U. L. REV. 905, 909 (2010) (quoting John Adams).

71. CUDDIHY, *supra* note 7, at 328–29.

century, and Sergeant William Hawkins's treatise in 1721 had specifically objected to general warrants empowering "a common Officer to arrest what Persons, and search what Houses he thinks fit."[72] Americans did not object to the use of general warrants until they were placed in the hands of customs officers.[73]

The writs of assistance added insult to injury by permitting customs officers to demand the assistance of citizens in conducting searches.[74] The right to call on the citizenry for assistance in the enforcement was not unique to the writs of assistance, but it prompted a public outrage in this context. If a constable was outmatched by the strength of either the offender he was to apprehend or the homeowner he was to search, he could call for a posse to assist him.[75] There were penalties for able-bodied men refusing to assist law enforcement officers, but typically members of the community very willingly came to the constable's aid.[76] Alexis de Tocqueville observed the eagerness of Americans to join in the hunt for an accused.[77] However, constables were enforcing laws that generally met with the approval of the public. By contrast, colonial customs agents, at least those working for the British Crown, were enforcing laws that were anything but popular and, when armed with Writs of Assistance, were able to demand that the public become complicitous in their offensive enforcement.

The abolition of general warrants did not, however, define the scope of authority for customs and revenue investigators, whose investigations were essential in a world prior to an income tax. It was, however, clear that general warrants were unlawful in early American Republic. American and English law had thoroughly repudiated general warrants by the time the colonies had separated from Great Britain but without a controversy in the colonies outside the customs context. The death knell for these warrants may well have been sounded in a case involving something other than

---

72. Davies, *supra* note 5, at 579, 629.

73. *See* discussion *supra* note 50 and accompanying text.

74. *See* William J. Stuntz, *The Substantive Origins of Criminal Procedure*, 105 YALE L.J. 393, 405 (2002).

75. *See* Steven J. Heyman, *Foundation of the Duty to Rescue*, 47 VAND. L. REV. 673, 689 (1994) (attributing the end of the requirement to assist police to the creation of nineteenth-century police departments charged with preventing crime).

76. *See* WILLIAM J. NOVAK, THE PEOPLE'S WELFARE: LAW AND REGULATION IN NINETEENTH-CENTURY AMERICA 57 (1996); Gautham Rao, *The Federal Posse Comitatus Doctrine: Slavery, Compulsion, and Statecraft in Mid-Nineteenth-Century America*, 26 LAW & HIST. REV. 1, 3 (2008) (observing that typically "[s]tates and localities exercised this power over persons with little apparent difficulty."). The meager state of law enforcement left a sort of citizen's veto in place. *See* LARRY D. KRAMER, THE PEOPLE THEMSELVES: POPULAR CONSTITUTIONALISM AND JUDICIAL REVIEW 26 (2004) (describing the necessity of citizen cooperation with a posse and the possibility of citizen interference with attempts to arrest).

77. ALEXIS DE TOCQUEVILLE, DEMOCRACY IN AMERICA 95–96 (Perennial Classics 2000).

customs enforcement, but this occurred in England, not the colonies. In the 1760s, an outlandish pamphleteer, who made a career out of maligning the king's ministers, waged a spectacular legal battle against general warrants.[78] General warrants were issued for the seizure of any papers revealing the authorship of *North Briton Number 45*, one installment of John Wilkes' weekly publication attacking King George III.[79] The warrants also called for the arrest of any persons suspected of authoring or publishing *Number 45*.[80] In a variety of suits filed by those, including Wilkes, who had been arrested or had their homes searched, the English courts provided precedent for the proposition, long espoused in the treatises, that general warrants were unlawful.[81] Following the Wilkes cases, colonial courts refused to issue general warrants to customs officers even though Parliament specifically authorized these warrants in the Townshend Acts.[82] The Fourth Amendment to the United States Constitution would then permanently memorialize this rejection of general warrants.

While the Framers were in accord in rejecting general warrants, they appear to have been of mixed minds about the sort of discretion customs officers should possess.[83] The broader principle comprehended by a ban on general warrants is still debated in Fourth Amendment cases today.[84] Early federal legislation imposed varying requirements on these officers seeking search warrants. The First Congress, admittedly prior to the adoption of the Fourth Amendment, authorized magistrates to issue warrants permitting customs agents to search any specific building the officer alleged contained

---

78. *See* ARTHUR H. CASH, JOHN WILKES: THE SCANDALOUS FATHER OF CIVIL LIBERTY 65–95 (2006).

79. *Id.* at 99–100; CUDDIHY, *supra* note 7, at 440–41.

80. CUDDIHY, *supra* note 7, at 440–41.

81. *Id.* at 444–46.

82. *See* Davies, *supra* note 5, at 702.

83. Davies, *supra* note 28, at 371 ("There is no indication in the historical record that the language of the Fourth Amendment was understood to alter the settled common-law standards for criminal arrest or search warrants.").

84. There is no shortage of efforts to analogize the discretion given to modern police to the discretion given customs officers under general warrants. *See, e.g.*, Jed Rubenfeld, *The End of Privacy*, 61 STAN. L. REV. 101, 132–60 (2008) (analogizing the problem with general warrants to use of undercover agents, detention of enemy combatants, and wiretapping); Barbara Salken, *The General Warrant of the Twentieth Century? A Fourth Amendment Solution to Unchecked Discretion to Arrest for Traffic Offenses*, 62 TEMP. L. REV. 221 (1989); Scott E. Sunby, *Protecting the Citizen "Whilst He is Quiet": Suspicionless Searches, "Special Needs" and General Warrants*, 74 MISS. L. J. 501 (2004) (arguing that intrusions lacking probable cause or a warrant justified under "special needs" exception is analogous to general warrant). In the late nineteenth century, analogies were similarly drawn to subpoenas for records of telegraphed communications. *See, e.g.*, *Ex Parte Brown*, 72 Mo. 83 (Mo. 1880). Andrew Taslitz has recently drawn an analogy between slave patrols in the antebellum South and general warrants. *See* ANDREW E. TASLITZ, RECONSTRUCTING THE FOURTH AMENDMENT: A HISTORY OF SEARCH AND SEIZURE, 1789–1868 12 (2006).

evidence of an import violation.[85] The Third Congress, following ratification of the Fourth Amendment, adopted Alexander Hamilton's Excise Act, which required revenue officers to provide magistrates with facts demonstrating probable cause to believe evidence of a violation could be discovered in the place described.[86] The First Congress's scheme did little to constrain the discretion of customs agents. It merely placed an administrative burden on them in the form of necessary paperwork. Much like victims in criminal cases, customs officers' allegations were sufficient for warrants. The Third Congress' scheme placed a genuine limitation on revenue agents. A disinterested magistrate had to agree with the (very) interested agent that probable cause existed. This latter scheme looked very much like the modern warrant standard under which a neutral and detached decision maker evaluates the basis of the officer's suspicions.

Under either scheme, customs and revenue officers had discretion that far eclipsed that of ordinary law enforcement officers. The mechanism for obtaining a warrant tells only part of the story. Much like in modern times, warrants were not always required. Each of these Congressional schemes also allowed officers to engage in warrantless seizures—customs agents could search ships without a warrant and revenue officers could search registered distillers without suspicion.[87] Beyond that, Congressional legislation and Marshall Court opinions made suits against customs officers for trespass extraordinarily difficult to win.[88]

Ironically, Americans, who had just fought a war of independence in large part over customs enforcement, conferred remarkably more authority on customs and revenue officers than they would contemplate giving officers enforcing criminal laws. Probable cause, as we understand it in modern terms, was sufficient for these searches, in some cases more than sufficient. Customs officers were permitted to seek authorization for searches and seizures on the basis of their investigations, or sometimes act on their own as a result of their investigations. They were financially rewarded when their suspicions were correct. Neither financial incentives nor legal standards equipped officers enforcing criminal laws with such motivation or discretion. The realities of urban life—and new types of criminal laws—would confer similar legal powers on police officers.

---

85. *See* CUDDIHY, supra note 7, at 757 (2009) (observing that under the Federal Collections Act of 1789, magistrates were given no discretion to refuse a customs officer's request for a warrant).

86. *Id.* at 757 (observing that Alexander Hamilton's Excise Act of 1791 required "reasonable cause of suspicion to be made out to the satisfaction of . . . [a] judge or justice.").

87. Amar, *supra* note 42, at 766; Cloud, *supra* note 14, at 1743 n.127.

88. *See* Arcila, *supra* note 65, at 420–21 (describing how early American customs statutes operated to limit access to remedies for unreasonable searches); Fallon & Meltzer, *supra* note 65, at 1783 n.279.

II. NINETEENTH-CENTURY PROHIBITION FASHIONED POLICE-FRIENDLY SEARCH STANDARD

As modern police departments were created, ordinary officers were given incentives—though not as overtly pecuniary as those given to customs officers—to investigate crime aggressively. With the development of full-time police forces, law enforcement became a career rather than a part-time obligation. Successful performance became a basis for retention and promotion.[89] Conferring power on these officers to conduct searches or make arrests on the basis of information discovered they discovered made sense, as their investigations were only useful if they could supplant, or at least supplement, victims' complaints.

Good record keeping in New York City and the abundance of secondary materials on its history make the city a good starting point when examining nineteenth-century changes in criminal procedure. When the City of New York created a force of career officers to suppress riots and investigate and prevent crimes, it increased the amount of manpower dedicated to law enforcement, developed a military-style hierarchy, and perhaps even increased the social standing of those responsible for policing the City.[90] The creation of the new force had changed the incentives for police officers. There were political motivations for those at the top of the hierarchy to at least appear to be suppressing crime, and those lower in the hierarchy had an interest either in climbing the ladder or simply retaining a better-than-average-paying job in the mid-nineteenth century.[91] The legal standards that had inhibited eighteenth-century constables did not, however, immediately change with the adoption of this new force. While the state legislature had authorized the force's creation a year earlier, it had done nothing to modify the search and seizure standards that had made police investigations legally irrelevant.[92] Because of public hostility to police departments, the legislature's abstinence on this issue is hardly surprising. Indeed, the creation of professional police departments had been thwarted

89. *See* Oliver, *supra* note 57 at 459–60 (describing rise of career officers).

90. *See* LISA KELLER, TRIUMPH OF ORDER: DEMOCRACY & PUBLIC SPACE IN NEW YORK AND LONDON 163 (2009); Davies, *supra* note 5, at 641. The new rules for the New York Municipal Police made it clear that officers were to be more proactive than their predecessor constables and watchmen. CITY OF NEW YORK, RULES AND REGULATIONS FOR THE GENERAL GOVERNMENT OF THE POLICE DEPARTMENT OF THE CITY OF NEW YORK 25 (1848) [hereinafter NEW YORK, RULES AND REGULATIONS] (stating that the "prevention of crime [is] the most important object" of the officer). Certainly the social standing of higher-ranking police officers in the early twentieth century exceeded that of any law enforcement officer in the eighteenth century. For example, Police Commissioners Teddy Roosevelt and Arthur Woods were both Harvard graduates and were definitely in the upper echelon of New York society.

91. *See* Oliver, *supra* note 57, at 459–60.

92. *See* RICHARDSON, *supra* note 54, at 51.

for nearly a decade by concerns about maintaining a military-style force with broad police powers.[93]

For nearly a decade after the creation of the Municipal Police Force, the law did not permit an officer to apply for a search warrant; crime victims alone retained this prerogative. The needs of law enforcement may have been a contributing factor to the willingness of New Yorkers to accept police-initiated searches, but the timing of the new standard demonstrates that another factor was far more substantial. Victimless crimes—to the extent they existed in the mid-nineteenth century—were neither regularly investigated nor prosecuted. It would take the Temperance Movement to whip up support for a new search procedure.

As the Temperance Movement shifted from moral suasion to successful advocacy of legislation in the mid-nineteenth century, it needed a mechanism to ensure compliance with state-wide prohibitory laws.[94] With a few rare exceptions in the mid-nineteenth century, searches outside the customs context could occur only if someone could swear that a crime had actually occurred.[95] The advocates of Prohibition did not seek an opportunity to conduct a search for an obscure crime that was rarely prosecuted. They sought a mechanism to search for the most commonly committed "crime" of their era, a mechanism that would have to be capable of frequent exercise if Prohibition were to have a chance of success. The drafters—and certainly the opponents—of this new law recognized that they were reshaping the rules relating to searches in ordinary criminal cases.[96] And in response to this new power, courts crafted a new mechanism for limiting police officers: the exclusionary rule, which would quickly be cabined to searches for liquor that lacked the requisite formalities.

Prohibition introduced New York, and many other states, to a warrant application process in ordinary criminal cases that did not require a victim's complaint. For as much as New York City owned the nineteenth century

---

93. *See* BURROWS & WALLACE, *supra* note 4 at 636–38. *See generally* William S. Fields & David T. Hardy, *The Third Amendment and the Issue of Maintenance of Standing Armies: A Legal History*, 35 AM. J. LEGAL HIST. 393 (1991).

94. JOHN A. KROUT, THE ORIGINS OF PROHIBITION 266 (1925); LORI D. GINZBERG, WOMEN IN THE WORK OF BENEVOLENCE: MORALITY, POLITICS, AND CLASS IN THE NINETEENTH-CENTURY UNITED STATES 98–132 (1990); Jed Dannenbaum, *The Origins of Temperance Activism and Militancy Among American Women*, 15 J. Soc. Hist. 235, 239–40 (1981) (explaining the Temperance Movement's shift from moral suasion to prohibition in terms of gender).

95. *See supra* discussion note 56 and accompanying text.

96. Efforts at liquor enforcement obviously required enhanced investigatory powers, which benefited the advocates of greater police authority. The relationship between Prohibitionists and advocates of stronger police powers was strained, however, as even the fledging policing organizations of the first half of the nineteenth century had recognized the perils of enforcing limits on alcohol. *See* ROBERT L. HAMPEL, TEMPERANCE AND PROHIBITION IN MASSACHUSETTS, 1813–1852 (1982).

(just as Boston and Philadelphia had owned the eighteenth century), the nineteenth-century Prohibition movement did not have its origins anywhere near Manhattan.[97] Its origins lie in a much smaller town whose historical significance is often underappreciated: Portland, Maine. Of course, Portland does seem an unlikely location to have spawned a national social reform or legal change. Maine was then, as it is now, a relatively unpopulated state, with most of its citizens residing in southern coastal towns. Portland itself was a mid-sized port, comparable to New Haven, Salem, and Charleston. Its status as an import and export hub was a blessing and a curse to the town of 20,000 in the antebellum era.[98] While, trade benefited the city, drunken sailors created a market for the ready flow of cheap rum. However, alcohol use in this town was not unique to transient sailors; over 300 bars and taverns operated within the city limits, some serving alcohol out of open troughs.[99] Minors as well as adults were intoxicated on Portland's streets at all hours of the day and night.[100]

It was not the character of the town, but rather the determination and single-minded devotion of one of its residents, that made Portland the home of the American Prohibition movement. Neal Dow was a Quaker far less passive than one might imagine for a man of that religious sect. He owned a tannery he had inherited from his father, was a leader in his local fire department, and had a reputation for being a firebrand orator prone to

97. Philadelphia had, of course, been the revolutionary capital and became the nation's financial capital in the eighteenth century, a status it maintained well into the nineteenth century. ROBERT E. WRIGHT, THE FIRST WALL STREET: CHESTNUT STREET, PHILADELPHIA, AND THE BIRTH OF AMERICAN FINANCE 11 (2005). Boston began the eighteenth century as America's premier city. *See* EVARTS BOUTELL GREENE, PROVINCIAL AMERICA, 1690–1740, *in* THE AMERICAN NATION: A HISTORY 244 (Albert Bushwell Hart ed., Classic Reprints 1964) (1905) ("during the first half of the eighteenth century Boston held its place as the most considerable centre of population and trade on the continent."). Both Philadelphia and New York had larger populations than Boston by outbreak of the American Revolution, but Boston had a political significance that neither of the larger cities could rival. The oft-held perception of New York as America's most important city did not emerge until the second half on the nineteenth century. *See* ROBERT A.M. STERN, THOMAS MELLINS & DAVID FISHMAN, NEW YORK 1880: ARCHITECTURE AND URBANISM IN THE GUILDED AGE 15 (1999) ("In the eyes of the so-called civilized world, and especially those in major European capitals, post-Civil War New York was only just beginning to come into focus as America's representative city."); DAVID MCCULLOUGH, GREAT BRIDGE: THE EPIC STORY OF THE BUILDING OF THE BROOKLYN BRIDGE 121 (1972) (New York "was the undisputed center of the new America that had been emerging since the [civil] war.").

98. CAMPBELL GIBSON, POPULATION DIVISION, U. S. BUREAU OF THE CENSUS, POPULATION DIVISION WORKING PAPER NO. 27, 1998, POPULATION OF THE 100 LARGEST CITIES AND OTHER URBAN PLACES IN THE UNITED STATES: 1790 TO 1990, Table B, *available at* http://www.census.gov/population/www/documentation/twps0027/twps0027.html (last visited Apr. 2, 2011).

99. NEAL DOW, THE REMINISCENCES OF NEAL DOW: RECOLLECTIONS OF EIGHTY YEARS 153–80 (1898).

100. *Id.* at 169.

lobbing personal attacks against his opponents.[101] Long a temperance advocate of powerful constitution, very early in his adult life he was heard to object to the excesses of alcohol use, particularly rum, in his city.[102] Inspired by a failed effort at statewide prohibition in Massachusetts in the 1830s, he embarked on a tireless campaign to create a criminal penalty for the sale, manufacture, or possession of alcoholic beverages in Maine in the 1840s.[103]

In form, Dow's idea was unusual; in substance, it was revolutionary. There were few victimless crimes on the books in the first half of the nineteenth century, and those that did exist were rarely enforced.[104] Their enforcement was so infrequent that no one had given much thought to the mechanism used to obtain a warrant to search for pornography, for instance.[105] Dow had not, however, proposed creating a run-of-the-mill victimless crime: he had targeted alcohol. Since the colonial era, local licensing laws had regulated alcohol, which ensured lenient liquor laws in towns in which demand for alcohol was high.[106] Port cities and rural

---

101. *See* FRANK L. BYRNE, PROPHET OF PROHIBITION: NEAL DOW AND HIS CRUSADE 9–16 (1961).

102. *Id.* at 12–24.

103. *Id.* at 24. General James Appleton made the attempt in Massachusetts, and his family would insist that he had not received his due for the later success in Maine. *See* D. F. APPLETON, THE ORIGIN OF THE MAINE LAW AND OF PROHIBITORY LEGISLATION, WITH A BRIEF MEMOIR OF JAMES APPLETON (1886).

104. At the turn of the nineteenth century, American authorities recognized the legitimacy of search warrants only to recover stolen goods. *See* OLIVER L. BARBOUR, A TREATISE ON THE CRIMINAL LAW AND CRIMINAL COURTS OF THE STATE OF NEW YORK 499 (2d ed. 1852); *In re* Special Investigations No. 228, 458 A.2d 820, 831 (Md. Ct. Spec. App. 1983) ("the common law of England and of Maryland recognized the search warrant for stolen goods, but no other search warrant."); Amar, *supra* note 42 at 765 (describing that "common law search warrants . . . were solely for stolen goods."). *But see* A. OAKEY HALL, A REVIEW OF THE WEBSTER CASE BY A MEMBER OF THE NEW-YORK BAR (New York, J.S. Redfield 1850) (rare case in which a search warrant was authorized without statutory authority for the search of a home for clothes which a witness claimed the culprit wore). Some early American statutes permitted searches for smuggled items or dangerous items such as gunpowder or diseased or infected items. *Id.* Treatise writer Joel Bishop recognized in 1880 that search warrants were most commonly issued for stolen goods although warrants to discover lottery tickets, intoxicating liquors, and gaming implements were beginning to be issued as new statutes created victimless crimes. 1 JOEL PRENTISS BISHOP, CRIMINAL PROCEDURE; OR, COMMENTARIES ON THE LAW OF PLEADING AND EVIDENCE AND THE PRACTICE IN CRIMINAL CASES 145 (3d ed. 1880).

105. *See* ME. REV. STAT. tit. XII, Ch. 160, § 19 (1841). Even though state statutes in Maine authorized a search for pornography, for instance, Governor Dana observed in 1850 that the only victimless crime investigated using a search warrant was unlawful possession of gunpowder. There were very rare exceptions. 13 THE MONTHLY LAW REPORTER 208–09 (Stephen H. Phillips, ed., Boston, Charles C. Little & James Brown 1851) [hereinafter MONTHLY LAW].

106. *See generally* KROUT, *supra* note 94 (tracing history of prohibition from colonial

villages under the new law would be treated alike; neither would be allowed to permit any sales of alcohol. This was an attempt at a social revolution through the criminal justice system. For somewhat obvious reasons, prohibition would become the most prosecuted victimless crime of the nineteenth century.[107]

Maine adopted Dow's first prohibitory bill in 1846, which was largely ineffective. Because it did noting to authorize searches for liquor, the law failed to put a meaningful dent in the amount of alcohol in the state, even by its proponents' estimates.[108] Witnesses alleging violations of liquor laws—often informants paid by Temperance Men—would testify to observing sales, but were seldom believed.[109] Prosecutions frequently suffered from a lack of physical evidence, as existing search and seizure doctrines did not permit searches for illegal alcohol.[110] There were no victims who could complain of an injury from a violation of the liquor law.

Dow would therefore return to the legislature in 1849 with a proposal to permit a search for evidence of this victimless crime.[111] Under the bill, any three persons could appear before a magistrate, allege that they had probable cause to believe liquor was in the location specified in the complaint, and obtain a warrant.[112] This was, of course, essentially the procedure in Maine, as in all early American states, for obtaining a search warrant to recover stolen goods.[113] There were some differences in the requirements for a search warrant, depending on whether an applicant wanted to search for alcohol or stolen goods. Dow's proposal required three complainants, while a search warrant for stolen goods could be obtained by

---

era to enactment of Maine Law).

107. One piece of data confirming this conclusion can be found in a late nineteenth-century digest. The search and seizure entry refers the reader to the section on intoxicating liquors. ALBERT R. SAVAGE, AN INDEX-DIGEST OF THE REPORTS OF CASES DECIDED BY THE SUPREME JUDICIAL COURT OF MAINE (1897). There were victimless crimes in the mid-nineteenth century, all of which depended on searches and seizures for prosecution. *See, e.g.*, BENJAMIN KINGSBURY, JR., THE JUSTICE OF THE PEACE: DESIGNED TO BE A GUIDE TO JUSTICES OF THE PEACE, FOR THE STATE OF MAINE 180 (Portland, Sanborn & Carter 1852) (describing types of items that could be sought under search warrant); ME. REV. STAT. tit. III, ch. 34, § 5 (1841) (permitting searches for improperly stored gunpowder); *id.* tit. XII, ch. 160, § 18 (allowing search warrant to discover young women in bawdy houses); *id.* § 20 (permitting warrants for obscene publications). Yet the digest entry for search and seizure notes that all of the cases decided in Maine on this topic have been considered in the context of intoxicating liquors.

108. Neal Dow himself recognized that without the search mechanism, his efforts would have been doomed to failure. *See generally Prohibitory Laws of Maine*, N.Y. TIMES, Feb. 3, 1896, at 5.

109. BYRNE, *supra* note 101, at 39.

110. *Id.* at 42.

111. *Id.* at 42–43.

112. *The New Liquor Law*, KENNEBEC JOURNAL, Aug. 23, 1849, at 3.

113. *See* notes 9–12 and accompanying text.

only one complainant. A complainant alleging stolen goods had to be a victim of the crime, while there were obviously no victims of the prohibitory laws. Each type of warrant, however, required only the complainant's allegation of his suspicions. Applicants for search warrants under Dow's proposal, just as applicants for search warrants to recover stolen goods, were not required to explain the basis of their suspicion.

Both houses of the legislature passed Dow's bill, but not enthusiastically. Searches for alcohol threatened a new degree of government intrusion. While a search for a stolen item could be initiated only if a victim identified missing property, presumably located only in a single location, a complainant could contend that liquor was housed in any number of locations. There was also concern about the character of applicants who would seek warrants for illegal liquor, for the same witnesses Dow and others had hired to bear witness against their neighbors under the old law were expected to appear as complainants under this new law.[114]

Many members of the legislature voted against the measure. However, others, opposed prohibition entirely or feared expanding the government's authority to search, voted for the bill because they were from pro-temperance districts. Legislators appear to have struck a deal with the outgoing Governor John Dana to veto the bill if it passed.[115]

When the legislature passed the bill, Governor Dana issued a preliminary statement summarily expressing his concern about the search and seizure provision.[116] Months later, he would issue a remarkably thorough veto message to the legislature.[117] Near the end of his life, Neal Dow paid this document a strong compliment, writing in his memoirs that "[f]rom that day to this nothing has been urged against Prohibition that was not expressed or implied in what Governor Dana had to say nearly half a century ago."[118] It could certainly be argued that Dow, a man with no wavering belief in the righteousness of his cause, was noting the lack of arguments that could be made against his reform. Far more likely, he was paying a genuine compliment to the thoroughness of a deceased and respected adversary. Dow also noted in his memoirs that Dana was "a man

---

114. *See* BYRNE, *supra* note 101, at 42 (observing Dow's difficulty in finding credible witnesses to liquor sales).

115. This inference is supported by the fact that a similar development occurred with the passage of the liquor law of 1851, which was successful. Several members of the legislature who voted for the bill counseled then-Governor John Hubbard to veto it, noting that they could not have voted for it and retained their seats. They advised him to follow the course of his predecessor. *See* DOW, *supra* note 99, at 340–43. Neal Dow also observed that Governor Dana had taken the "counsel of some of the leaders in his party" in vetoing the bill. *Id.* at 320.

116. *Closing Proceedings of the Legislature*, KENNEBEC JOURNAL, Aug. 23, 1849, at 3.

117. MONTHLY LAW, *supra* note 105, at 205–13.

118. DOW, *supra* note 99, at 319.

of ability and influence, and justly entitled to leadership among his political associates."[119]

In the portion of the message dealing with the search provision, Governor Dana observed that common law protections against unreasonable searches were inapplicable to searches for evidence of this crime.[120] He acknowledged that searches for other items could be initiated by a mere complaint, but observed that most frequently searches were to recover stolen goods. In order to initiate this most common search,

> there must be a pre-existing fact, not merely suspected, but *known* to the complainant, to wit, the loss of the goods; and when such a fact exists, the person suffering the loss, in instituting search, will give to it only that direction which the circumstances may indicate, as most likely to result in the recovery of his property.[121]

With no victim to swear to an injury, and no specific goods to search for, Governor Dana contended that there was no limit on the number of searches that could be authorized and no end point to a search for liquor.[122] The governor recognized that the Maine Legislature had previously authorized searches for some victimless crimes—crimes for which no one could swear to an injury. Searches were permitted, for instance, for pornography,[123] prostitutes,[124] gambling instruments,[125] and illegally stored gunpowder.[126] Of these, only gunpowder searches were conducted with any degree of frequency, and this was likely due to the extraordinary number of gunpowder mills that had cropped up shortly before Maine's statehood.[127] Governor Dana noted that, unlike in the case of alcohol searches, there was "no danger of general abuse" of the gunpowder warrant, as "the number is

---

119. *Id.* at 321.

120. MONTHLY LAW, *supra* note 105, at 208–09.

121. *Id.* at 208.

122. *Id.* at 208–09

123. ME. REV. STAT., tit XII, ch. 160, § 20 (1841).

124. *Id.* § 18.

125. *Id.* § 39.

126. One of the earliest statutes of Maine provided that a search warrant could be obtained by a selectman of the town to investigate the possibility that gunpowder was being stored contrary to the regulations of the town. An Act for the Prevention of Damage by Fire, and the Safe Keeping of Gun Powder, ch. 25, § 5, 1821 Me. Laws 112,114 (1821); JOHN MAURICE O'BRIEN, THE POWERS AND DUTIES OF THE TOWN OFFICER, AS CONTAINED IN THE STATUTES OF MAINE 261 (Hallowell, Glazier, Masters & Smith 4th ed. & Co. 1840). Statutes regulating the possession of gunpowder in early American states were somewhat common. *See* Saul Cornell & Nathan DeDino, *A Well-Regulated Right: The Early American Origins of Gun Control*, 73 FORDHAM L. REV. 487, 510–12 (2004).

127. *See* MAURICE M. WHITTEN, THE GUNPOWDER MILLS OF MAINE 3 (1990) (indicating that around 1820, entrepreneurs in Maine sought to establish mills in the new state, where there had previously been none).

small to whom the suspicion could possibly attach, of violating the law, which regulates the keeping of gunpowder."[128]

Governor Dana's message reveals something very interesting about the protections eighteenth- and early nineteenth-century criminal procedure provided against unreasonable searches and seizures. The common law limitations on warrants had ensured that searches would be relatively *rare*, not that they would necessarily be *accurate*. Neal Dow had proposed greatly enlarging the role of the state by authorizing searches of homes to discover evidence of a frequently violated law. Searches for alcohol under his new prohibitory law would not be rare. To get his bill enacted, Dow would have to convince the legislature and the governor that he had discovered a mechanism to enhance the accuracy of searches.

A year after Governor Dana's veto, Neal Dow returned to the legislature with a bill that would not only be enacted in Maine, but would be adopted in several American jurisdictions. This bill, like Dow's previous bill, permitted magistrates to issue warrants to search for liquor when three voters alleged they had probable cause to believe alcohol could be located in the specified location.[129] The bill, however, forbade a search of a dwelling house unless one of the three complainants swore that he witnessed an alcohol sale out of the house.[130] Like Dow's previous attempt, this bill passed both houses of the legislature, and Dana's successor, John Hubbard, signed it into law on June 2, 1851.[131] This law would forever link the state with the prohibition movement, as around the world, prohibitionists would advocate adopting the "Maine Law."[132]

The provisions of the new law obviously required a magistrate to review a complaint containing facts supporting the affiant's conclusion that a crime had been committed. This was of course the process Hale, Hawkins, and Blackstone had prescribed for all search warrants, but which

---

128. MONTHLY LAW, *supra* note 105, at 208. Governor Dana's analysis here is not unlike the justification for warrantless searches of closely regulated businesses. *See* 5 WAYNE R. LAFAVE, SEARCH AND SEIZURE: A TREATISE ON THE FOURTH AMENDMENT 37–95 (4th ed. 2004).

129. Voting requirements in Maine were not particularly stringent in the mid-nineteenth century. All males, including African Americans, who were neither aliens nor paupers, and who had established a residence in the state for at least three months, were entitled to vote. *See* Opinion of the Supreme Judicial Court, 44 Me. 507 (1857) (responding to question posed to the court by the state senate).

130. An Act for the Suppression of Drinking Houses and Tippling Shops, 1851 Me. Laws 210, 214–15, Me. Rev. Stat. ch. 211, § 11 (1851).

131. *Id.* at 215.

132. *See* DOCUMENTARY HISTORY OF THE MAINE LAW: COMPRISING THE ORIGINAL MAINE LAW, THE NEW-YORK PROHIBITORY LIQUOR LAW, LEGISLATIVE DEBATES, ARGUMENTS, JUDICIAL DECISIONS, STATISTICS, IMPORTANT CORRESPONDENCE; "INQUISITION" AND PROHIBITION VERSUS "FREEDOM" AND ANTI-PROHIBITION 85–86 (New York, Hall & Brother 1855) [hereinafter DOCUMENTARY HISTORY].

was rarely, if ever, followed in practice.[133] For many members of the legislature, this standard not only provided a mechanism to prevent false searches, it also may have seemed a comfortable resort to a procedure deeply rooted in Anglo-American history. The statute, that is, may have evoked a sense of nostalgia for a past that never existed.

Neal Dow was widely (and falsely) credited as author of this new search provision,[134] which contemporaries recognized as fundamentally changing the law. Reformers, including Lyman Beecher, Horace Mann, and Sam Houston, hailed the passage of the new law.[135] Attracting such national attention were Dow's tireless self-promotion and the only substantial change from Maine's 1846 statewide prohibitory law, the new enforcement mechanism specified in the act's warrant section. Fellow prohibitionist John Marsh dubbed Neal Dow the "Napoleon of Temperance" and hailed the search and seizure provision for making prohibition a reality. Dow, Marsh wrote, had "brought into the battle-field every officer of the State, . . . turned its whole artillery against the rum-fortifications, and in less than six months, . . . swept every distillery and brew-house, hotel-bar, splendid saloon and vile groggery clean from the State."[136]

It would have been surprising if Dow had developed a standard that would have been so familiar to lawyers. Dow, a tanner by trade, never studied law, although it was his dream.[137] His father had great disdain for lawyers and insisted that his son not attend college,.[138] In his memoirs, Neal Dow noted that he received some "technical" assistance in writing the Maine Law from Edward Fox, a prominent Portland lawyer who later would be appointed a federal district judge by Andrew Johnson.[139]

---

133. *See* sources cited *supra* note 21.

134. *See e.g.* BYRNE, *supra* note 101, at 45 ("Dow's greatest innovation was the provision for search and seizure."); HENRY S. CLUBB, THE MAINE LIQUOR LAW: ITS ORIGIN, HISTORY, AND RESULTS, INCLUDING A LIFE OF HON. NEAL DOW 23 (New York, Fowler & Wells 1856) ("Still persevering, Neal Dow again appeared in the Hall of Representatives in August, 1850, with a bill of his own drafting, subsequently known as the 'Maine Law.'"); ALLAN LEVINSKY, A SHORT HISTORY OF PORTLAND 79 (2007).

135. *See* BYRNE, *supra* note 101, at 49, 141.

136. *Id.* at 48.

137. Though he had no legal training, Dow did once appear as counsel to defend a woman who was charged with horsewhipping a rum-shop keeper for selling liquor to her husband. The woman requested that Dow be permitted to act as her lawyer, and notwithstanding his lack of training in the law, the judge permitted him to do so. The jury found her guilty but recommended mercy, and she was required to pay, as Dow later recalled, "a slight fine" which he paid. DOW, *supra* note 99, at 99.

138. *See id.* at 56–58; IX S.M. WATSON, THE MAINE HISTORICAL AND GENEALOGICAL RECORDER 1884–1898 226 (1973). It was boasted in his father's obituary that he had only once resorted to the legal system in a suit to successfully recover a debt against the advice of his lawyer. Death Notice of Josiah Dow, Fox Family Scrapbooks, Vol. 3, Collection 849, Maine Historical Society (describing father's sole resort to the law).

139. *See* DOW, *supra* note 99, at 334–35 ("Having completed [the bill] to my own

Fox's assistance was far more substantial than Dow would ever publicly acknowledge. A graduate of Harvard College and Harvard Law School, Fox was extremely well regarded as a scholarly and knowledgeable attorney.[140] He was, therefore, likely either already familiar with Blackstone's description of the process for seeking a search warrant or became familiar with this description. Even if he never handled a criminal case, he would have had ready access to a volume with this description of the warrant application process. One of the earliest American versions of Blackstone's *Commentaries on the Laws of England* was published in Portland, and early Maine manuals for justices of the peace reiterated this description, which was, as a practical matter, never followed.[141] Blackstone's *Commentaries* were a staple in the library of mid-nineteenth-century lawyers, and Fox was surely no exception.[142] He was also much more likely than Dow to be familiar with the similar procedure required to obtain a warrant under the Excise Act of 1791. These facts alone would suggest that Fox was the more likely author of the new search and seizure provision.

Open letters, published in Portland newspapers, between Neal Dow and his cousin, John Neal, confirm Fox's role in creating this provision. John, also a Portland lawyer, had initially been a supporter of his cousin's efforts to enact the Maine Law. However, a feud developed between the two, largely over a client of John's, a notorious Portland prostitute named Margaret Landigren, alias "Kitty Kentuck." She was convicted of violating the new liquor law—a charge John Neal believed to be false.[143] When John personally put up bond for her appeal, Neal Dow alleged that his cousin was having an affair, or at least a series of commercial transactions, with

---

satisfaction, I submitted it to Edward Fox . . . . He suggested a few changes, principally on technical points, which I accepted.").

140. For biographical information on Edward Fox, *see* N. M. FOX, A HISTORY OF THAT PART OF THE FOX FAMILY DESCENDED FROM THOMAS FOX OF CAMBRIDGE, MASS. 47 (St. Joseph, Mo., Union Printing 1899); HERBERT T. SILSBY, II, MEMORABLE JUSTICES AND LAWYERS OF MAINE 188–91 (2006); WILLIAM WILLIS, A HISTORY OF THE LAW, THE COURTS, AND THE LAWYERS OF MAINE, FROM ITS FIRST COLONIZATION TO THE EARLY PART OF THE PRESENT CENTURY, at iv (Portland, Bailey & Noyes 1863).

141. *See* SIR WILLIAM. BLACKSTONE, KNT., COMMENTARIES ON THE LAWS OF ENGLAND (Portland, Thomas B. Wait 1807); JEREMIAH PERLEY, THE MAINE JUSTICE 75–76 (Hallowell, Goodale, Glazier 1823) (stating standard from Blackstone). I am grateful to Chris Livesay, who allowed me to spend a day going through these and other original nineteenth-century treatises he has collected in his Brunswick, Maine law office.

142. *See* LAWRENCE M. FRIEDMAN, A HISTORY OF AMERICAN LAW 463 n.1 (3d ed. 2005) (describing Lincoln's reliance on Blackstone); ROBERT STEVENS, LAW SCHOOL: LEGAL EDUCATION IN AMERICA FROM THE 1850S TO THE 1980S 47–48 nn.36–37 (1983) (observing that Blackstone's *Commentaries* provided the curriculum for early American law schools).

143. *See* Matthew J. Baker, *The Saga of Portland's Unsinkable, Irish Kitty Kentuck*, PORTLAND MONTHLY MAGAZINE, Dec. 1996, at 24, 25–27; JAMES MUNDY, HARD TIMES, HARD MEN: MAINE AND THE IRISH 1830–1860 90–91 (1990).

Kitty.[144] Angry letters between the two contained a variety of allegations, one of which Neal Dow never refuted.[145] John Neal alleged that his cousin was accepting accolades from all over the globe for drafting a search and seizure law everyone in the Portland community knew was drafted by Edward Fox.[146]

This new standard was indeed groundbreaking and, for prohibitionists, certainly worthy of the praise it received even if the wrong person was lauded. Fox's work had produced a standard that required a very specific type of proof to authorize a search. This new standard, however, took a step toward the modern probable cause standard in expressly requiring consideration of the facts supporting a complainant's accusations.

Prohibitionists turned to the legislature two years later to amend the statute they had successfully passed. It was a creative effort to permit liquor searches whenever three complainants swore that they had probable cause to believe alcohol could be discovered in the search requested and jettisoned the requirement that one of the complainants observe and testify to a liquor sale on the premises. Prohibitionists were attempting to install the original standard Neal Dow had proposed in 1849, which would have permitted a search whenever a complainant swore he had probable cause. Ironically, this effort would produce a rule that required a magistrate to review whatever facts a complainant offered in support of a search and determine whether sufficient suspicion existed to justify the search. The standard, in this generic form, could be applied to search (or arrest) warrants for anything, not just liquor. From this generic standard, it would be no great leap to permit officers to perform arrests when the facts available to them provided probable cause to believe a crime had occurred and the suspect had committed it.

The proposed amendment used vague language in an apparent attempt to dupe legislators into passing a law permitting a liquor search on the oath

144. *See* John Neal, *The Liquor Law of Maine – No. 2*, MAINE EXPOSITOR, Sept. 7, 1853, at 1; JOHN NEAL, WANDERING RECOLLECTIONS OF A SOMEWHAT BUSY LIFE 370–72 (1869).

145. *See* Neal Dow, *John Neal and the Liquor Law of Maine*, MAINE EXPOSITOR, Sept. 14, 1853, at 1 (reprinting article from the newspaper *State of Maine*); John Neal, *Mr. Neal's Reply*, MAINE EXPOSITOR, Sept. 14, 1853, at 1.

146. John Neal stated that in drafting the Maine Law, Dow "had the help of a legal personage, for whom we profess to feel a sincere regard, in preparing the very portions which are most offensive and preposterous, and which mainly distinguish it from the old law. What those are, will be seen hereafter, as we proceed with the 'searching analysis' we have in our mind." John Neal, *The Liquor Law of Maine*, MAINE EXPOSITOR, Aug. 31, 1853, at 2. Given Dow's reference to Fox's "technical assistance," the reference is not difficult to decode, but subsequent writings from Neal would clarify any ambiguity. John Neal would quickly grow considerably less charitable toward Fox when he, one week later, specifically named him, noting that he was "the gentleman who ranks among one of the putative fathers of the Maine Liquor Law, and is rather disposed to glory in the co-partnership, though he thinks it too merciful." Neal, *supra* note 145.

of complainants that they possessed probable cause. Under the 1853 bill, three persons who were competent to be witnesses in civil cases were required to allege that alcohol could be discovered in the requested search.[147] A magistrate could not issue a warrant to search a dwelling unless he was convinced "by the testimony of witnesses upon oath, that there is reasonable ground for believing" that unlawfully possessed liquor was in the house.[148] The new bill imposed a hefty penalty for perjury—one year in the state penitentiary.[149] Magistrates were required to record the statements of these complainants, and the complainants were required to sign the transcriptions of their testimony.

Opponents of the new bill alleged that the vague language in this provision would permit a search warrant on the mere oath of a complainant that he had reasonable grounds for his belief. Supporters of the bill suggested that the vague language did not change the law and pointed to the severe perjury penalty.[150] The final version adopted by the legislature, and signed by Governor Hubbard on April 1, 1853, differed from the initial bill only in punishment for perjury—two years in the final bill.[151]

Temperance forces quickly tested the parameters of the new law, seeking warrants to search dwellings for liquor without providing any facts to support the complainants' conclusion that alcohol was indeed present. A number of decisions from the state's highest court concluded that complainants were required to provide facts supporting their suspicions, and when this factual support was lacking, the court arrested the judgment

---

147. With the exception of women, this did not substantially open up the pool of potential complainants given Maine's otherwise very liberal suffrage laws. Women, of course, played a substantial role in the Temperance Movement, so this provision may have been perceived to greatly enlarge the number of informants appearing before magistrates. *See generally* HOLLY BERKLEY FLETCHER, GENDER AND THE AMERICAN TEMPERANCE MOVEMENT OF THE NINETEENTH CENTURY (2008). One member of the Maine Legislature objected to permitting women and aliens to seek search warrants. Hon. Geo. M. Chase, Speech In Opposition to the Additional Bill for the Suppression of Drinking Houses and Tippling Shops (March 26, 1853), *in* MAINE EXPOSITOR, April 27, 1853, at 1.

148. *See* REPORT OF JOINT SELECT COMMITTEE ON SO MUCH OF THE ADDRESS OF THE GOVERNOR AS RELATES TO THE ACT FOR THE SUPPRESSION OF DRINKING HOUSES AND TIPPLING SHOPS, 23 DOCUMENTS PRINTED BY ORDER OF THE LEGISLATURE OF THE STATE OF MAINE 26 (1853) (reciting bill).

149. *Id.* at 27.

150. *Id.* at 4 (noting that to search a dwelling house, "evidence of witnesses [had to] be given in writing, on oath, filed with the magistrate, sufficient to show that there is good ground to believe that spirituous and intoxicating liquors are kept or deposited therein").

151. An Act in Addition to Chapter Two Hundred and Eleven of Eighteen Hundred and Fifty One, ch. 48, § 11, 1853 Maine Laws 51, 59. There do not appear to have been any convictions for perjury under the statute. There was one in Rhode Island under a similar statute, but this is the only one I have discovered reported either in the appellate reports, or newspapers, from the 1850s. *See Perjury*, MAINE EXPOSITOR, June 22, 1853, at 2.

against the defendant and returned his alcohol.[152] The court thus interpreted the amendment to the liquor law to have created a generic standard for warrant applications. So, while complainants were no longer required to observe, and testify to, a liquor sale to obtain a search warrant, they were required to testify to the facts they alleged provided reasonable grounds to believe alcohol present.

The remedy the court afforded for the violation was certainly novel. Arresting the judgment of a lower court because of a defect in the warrant effectively forbade a court to consider the fruit of an unlawful search.[153] These decisions appear to be the first American decisions based on the principle modern lawyers have come to know as the exclusionary rule.[154] Two decades later, the court would hold that the fruits of an officer's unlawful, warrantless search were admissible, retreating from the full implications of the new remedy it had fashioned.[155] Even this limited version of the exclusionary rule, however, represented a substantial innovation in the law as reliable evidence had always been admissible throughout Anglo-American history regardless of how it was discovered.[156]

---

152. *See, e.g.*, State v. Staples, 37 Me. 228, 230 (1854) (holding mere allegation of presence of alcohol insufficient for a warrant); State v. Spirituous Liquors, 39 Me. 262, 263 (1855) (holding the warrant was "fatally defective" because it was not signed by the witnesses).

153. The court had previously concluded that complaints seeking warrants that failed to allege that the alcohol was intended for sale in the town where it was housed were defective and the proceedings under them must be quashed. *See* State v. Spirituous Liquor, 33 Me. 527, 530 (1852). The exclusionary rule had thus been previously established in a case in which the pleadings in the complaint were inadequate. The cases, following the 1853 law, applied this remedy to a failure in the sufficiency of the proof supporting the allegations in the complaint. *See Staples*, 37 Me. at 229-30.

154. It is frequently assumed that the exclusion of illegally obtained evidence "first appeared in a cryptic statement in the 1886 decision *Boyd v. United States*, [and] did not fully emerge until the 1914 decision in *Weeks v. United States*." Davies, *supra* note 16, at 622–23 (citation omitted).

155. *See* State v. McCann, 61 Me. 116, 118 (1873) (holding conviction under liquor law will not be disturbed when evidence is unlawfully obtained by officer who acted without a warrant).

156. Amar, *supra* note 42 at 785–87; Davies, *supra* note 16 at 623–24 n.17 (citing Commonwealth v. Dana, 443 Mass. (2 Met.) 329 (1841)) ("The Massachusetts Supreme Court first upheld the constitutionality of the statute but nevertheless announced that it was contrary to common law to permit an inquiry into how evidence was obtained during the course of a trial, a rule that became known as the 'collateral issue' doctrine.") Federal courts well into the twentieth century would wrestle with a variety of justifications for the exclusionary rule. *See* Davies, *supra* note 16, at 624–25; Potter Stewart, *The Road to Mapp v. Ohio and Beyond: The Origins, Development and Future of the Excl[u]sionary Rule in Search-and-Seiz[u]re Cases*, 83 COLUM. L. REV. 1365, 1372–77 (1983). The judicially created remedy in the mid-nineteenth century—that returned the defendant's liquor and dismissed the conviction against him—can only be explained by an effort to make the defendant whole after having been unlawfully prosecuted for a crime that the courts found to

Much of the fossil record of modern criminal procedure can thus be found in the Supreme Judicial Court's interpretation of the nation's first prohibitory laws. With this 1853 modification to the liquor law, and its interpretation by the court, Maine had fashioned a standard for search warrants that would be familiar to a twenty-first century lawyer. Affiants were no longer permitted to provide what modern Supreme Court decisions describe as "bare bones" affidavits.[157] The actual practice of justices of the peace—at least in liquor cases—now conformed with Blackstone's description of a magistrate's role in reviewing requests for search warrants. For the first time, treatises in Maine contained forms for magistrates to record the facts supporting the allegations of complainants.[158] And suppression of evidence replaced tort suits as the mechanism for preventing at least a category of illegal searches. The historical roots of the exclusionary rule may, therefore, be greater than its critics—and even its proponents—have recognized.[159] Justice Potter Stewart observed in the *Columbia Law Review* that the "first case associated with the exclusionary rule is *Boyd v. United States*" from 1886.[160] However, the rule has a somewhat older lineage than that once one looks to state cases—in Maine, a form of the rule was developed in response to a new power to search for liquor. So, while it is certainly true that no form of the exclusionary rule existed during the colonial era or in the early years of the republic, a version of the exclusionary rule was fashioned contemporaneously with ordinary officers acquiring the discretion of customs officers.[161]

---

be of dubious legitimacy. *See* Robert Post, *Federalism, Positive Law, and the Emergence of the American Administrative State: Prohibition in the Taft Court Era*, 48 WM. & MARY L. REV. 1, 7 (2006) (observing that the courts in the twentieth century doubted the legitimacy of Prohibition).

157. *See e.g.*, United States v. Leon, 468 U.S. 897, 915, 923 n.24 (1984); Illinois v. Gates, 462 U.S. 213, 239 (1983).

158. BENJAMIN KINGSBURY, JR., THE JUSTICE OF THE PEACE: DESIGNED TO BE A GUIDE TO JUSTICES OF THE PEACE, FOR THE STATE OF MAINE 295 (Portland, Sanborn & Carter 1852).

159. *See e.g.*, Luke M. Milligan, *The Source-Centric Framework to the Exclusionary Rule*, 28 CARDOZO L. REV. 2739, 2747–56 (2007) (looking at Justice Holmes' opinion in *Silverthorne Lumber Co. v. United States*, 251 U.S. 385 (1920), to explain the modern exclusionary rule); Jerry E. Norton, *The Exclusionary Rule Reconsidered: Restoring the Status Quo Ante*, 33 WAKE FOREST L. REV. 261, 263–67 (1998) (describing the development of the exclusionary rule).

160. Potter Stewart, *The Road to* Mapp v. Ohio *and Beyond: The Origins, Development, and Future of the Exclusionary Rule in Search-and-Seizure Cases*, 83 COLUM. L. REV. 1365, 1372 (1983).

161. Thomas Davies has concluded that the United States Supreme Court "transferred [an] expanded concept of government illegality to the new law enforcement officer by ruling in 1914 in *Weeks* that a federal marshal's unlawful warrantless search of a residence violated the Fourth Amendment and, thus, was subject to the constitutional logic of nullity." Davies, *supra* note 16, at 625. The exclusionary rule, he argues, "arose contemporaneously with the modern conception of the modern law enforcement officer." *Id.* The process he describes

Versions of Maine's prohibitory law quickly spread through the country, winning acceptance in legislatures in every region of the country except the Southeast.[162] The year after Maine adopted the prohibitory law with its search and seizure provision, a bill containing very similar provisions was positively reported out of a committee of the New York Legislature.[163] In a preview of future events, Democratic Governor Horatio Seymour objected to the infringement of civil liberties in the Maine Law.[164] Later Seymour would oppose capital punishment and object to Abraham Lincoln's arbitrary arrests of those suspected of disloyalty.[165] One of his earliest public positions advancing a civil libertarian position, however, was his objection to the Maine Law.

Governor Seymour objected that the Maine Law permitted searches not previously allowed in ordinary criminal cases that worked a violation of the federal constitution.[166] Seymour's criticism demonstrated the most precise knowledge of search and seizure law of any of the objections to the nineteenth-century prohibitory laws. Searches had long been authorized on far less certainty that a crime had occurred, but not searches to reveal evidence of ordinary crimes. Customs searches had been permitted whenever customs or revenue officers had probable cause to believe goods had been unlawfully imported, or that required taxes had not been paid on them.[167] Throughout American history, however, customs and revenue

---

occurring at the federal level seems to have occurred at the state level, first in Maine, then in states adopting the exclusionary rule as a new type of crime greatly expanded the role of officers.

162. *See* WILLIAM BLACKWOOD & SONS, BLACKWOOD'S EDINBURGH MAGAZINE, at 211 (1867) (describing the thirteen states to adopt the Maine Law and the efforts to secure its adoption in all the states); STEWART MITCHELL, HORATIO SEYMOUR OF NEW YORK 154 (1938) ("One state after another played with the reform until Maine laws were being argued over almost everywhere.").

It has been assumed that the Prohibition movement was not successful beyond these regions because of the linkage between the Prohibition and Abolition movements. Prohibition, however, came very close to becoming law in at least parts of the South in the mid-nineteenth century. *See* Thomas H. Appleton, Jr., *"Moral Suasion Has Had It's Day": From Temperance to Prohibition in Antebellum Kentucky, in* JOHN DAVID SMITH & THOMAS H. APPLETON, JR. (eds.), A MYTHIC LAND APART: REASSESSING SOUTHERNERS AND THEIR HISTORY 19–42 (1997).

163. DOCUMENTARY HISTORY, *supra* note 132.

164. *See* MITCHELL, *supra* note 162, at 156.

165. *See* Governor Horatio Seymour, Annual Message, (January 7, 1863) *in* V MESSAGES OF THE GOVERNORS, COMPRISING EXECUTIVE COMMUNICATIONS TO THE LEGISLATURE AND OTHER PAPERS RELATING TO LEGISLATION FROM THE ORGANIZATION OF THE FIRST COLONIAL ASSEMBLY IN 1683 TO AND INCLUDING THE YEAR 1906 465 (Charles Lincoln ed., 1909) (describing opposition to death penalty); MITCHELL, *supra* note 162, at 267; *see generally* MARK E. NEELY, THE FATE OF LIBERTY: ABRAHAM LINCOLN AND CIVIL LIBERTIES (1991).

166. Governor Seymour's Message, N.Y. TIMES, May 7, 1852, at 3.

167. Thomas, *supra* note 1, at 1477, 1493.

officers had been regarded to have greater searching authority than ordinary police officers.[168] Giving every officer in America the power to search homes for alcohol that customs officers had to search warehouses for untaxed goods offended Seymour. The Governor recognized that the Maine Law was breaking down the separate system of criminal procedure by giving ordinary officers powers comparable to customs agents.[169] Even the supporters of the Maine Law recognized that the search and seizure provision had worked a change; they however applauded the change.

Notwithstanding Governor Seymour's objections, the Maine Law would be well received in his state—and a number of others. A stronger version of the law passed both houses of the New York Legislature in 1854. Under this version, a search was authorized if any two voters complained that unlawful alcohol was kept for sale in the county or town in which the complaint was made. Seymour vetoed this bill, offering in 1854 as one of his reasons that it effectively authorized general searches.[170] In contrast to his precision in his 1852 critique, this was sloppy. There had never been agreement on the principle that made general searches objectionable, but New York's version of the Maine Law certainly required as much specificity as any procedure for authorizing customs or revenue searches in

---

168. Perhaps recognizing an opening to expand the powers of federal officers, within a decade of the Maine Law's widespread adoption, Congress expanded the power of customs officials to seize the books and papers of merchants that could be used to demonstrate revenue and import violations. *See* S.B. EATON, SEIZING BOOKS AND PAPERS UNDER THE REVENUE LAWS 5 (1874). This law would be famously rejected in the landmark case of *Boyd v. United States*, 116 U.S. 616 (1886), forbidding the seizure of books or records that merely recorded evidence that a crime had occurred, a rule which endured until the realities of the administrative state required the capacity to examine such records. *See* William J. Stuntz, *The Substantive Origins of Criminal Procedure*, 105 YALE L.J. 393, 419–28 (1995).

169. Seymour specifically charged that the search warrants issued under the Maine Law would provide for the general warrants that the Fourth Amendment had forbidden. Of course, the Fourth Amendment's limitations did not apply to the states, but despite Supreme Court precedent clearly stating this, there was a widespread belief in the mid-nineteenth century that they did. *See* Jason Mazzone, *The Bill of Rights in Early State Courts*, 92 MINN. L. REV. 1, 35–37 (2007). Seymour's description of the warrants, authorized under this law as "general warrants," seems a sloppier criticism than one might expect from Seymour given the precision of his description of the state of search and seizure law in 1852. Governor Seymour's Message, *supra* note 166, at 3. There had been a wide range of thought on exactly what made general warrants problematic. *See* CUDDIHY, *supra* note 7, at 580–81. The requirements under the Maine Law for a search warrant satisfied even the original Virginia Constitution's very thorough objection to general warrants. *See* Davies, *supra* note 61, at 100.

170. MITCHELL, *supra* note 162, at 155–56; Governor Horatio Seymour, Veto of a Bill Entitled "An Act for Suppression of Intemperance," (March 31, 1854) *in* IV MESSAGES OF THE GOVERNORS, COMPRISING EXECUTIVE COMMUNICATIONS TO THE LEGISLATURE AND OTHER PAPERS RELATING TO LEGISLATION FROM THE ORGANIZATION OF THE FIRST COLONIAL ASSEMBLY IN 1683 TO AND INCLUDING THE YEAR 1906 755 (Charles Lincoln ed., 1909).

the early republic.[171] It also required more assurances of accuracy than were required for any application for a customs search. Two witnesses were required to swear that probable cause existed and at least one of them had to provide facts under oath supporting their conclusions.[172] There was, however, an analogy to the type of fear that general searches produced: widespread searches. The search provision of the Maine Law sought to discover something that many New Yorkers had in their possession, and intended to keep in their possession.

With the temperance lobby now solidly against him, Horatio Seymour, like John Dana in Maine before him, lost the subsequent election. Seymour's successor Myron Clark signed into law a version of the Maine Law slightly different from the one Seymour vetoed.[173] Under this version, any "credible person" could complain to a magistrate that alcohol was kept or deposited in violation of the law.[174] The complainant was required to provide in writing, under oath or on affirmation, "the facts and circumstances upon which such belief is founded."[175] The statute then expressly recognized the screening role that the magistrate was to play. A magistrate was to issue the search warrant only "if he [was] satisfied that there [was] probable cause for said belief."[176] Earlier versions of the Maine

171. Seymour contended that alcohol could not be particularly described but certainly the same was true for many things for which search warrants had been sought throughout Anglo-American history. Governor Seymour's Message, *supra* note 166, at 3. Money, for instance, was certainly fungible.

172. It is not clear what inspired the multiple complainant rule that began with the original version of the Maine Law. Multiple witnesses were of course required in treason prosecutions. *See* L.M. Hill, *The Two-Witness Rule in English Treason Trials: Some Comments on the Emergence of Procedural Law*, 12 AM. J. LEGAL HIST. 95, 95 (1968); John H. Wigmore, *Required Numbers of Witnesses: A Brief History of the Numerical System in England*, 15 HARV. L. REV. 83, 99 (1901).

173. Much like Henry Dutton in Connecticut, Myron Clark established himself as one of the chief proponents of the Maine Law and, like Dutton, this stance launched him into a brief stay in the Governor's Office. *See* discussion *supra* note 26 (discussing Dutton); *see also* WILLIAM E. GIENAPP, THE ORIGINS OF THE REPUBLICAN PARTY, 1852–1856 153 (1987); MYRON HOLLEY CLARK, THE MAINE LAW: SPEECH OF HON. MYRON H. CLARK, 29TH DISTRICT, ON THE BILL FOR THE SUPPRESSION OF INTEMPERANCE IN THE SENATE, MARCH 3D 1854.

Thomas Davies and Fabio Arcila disagree as to whether colonial and early American magistrates screened the basis a complainant offered for believing that stolen goods would be found in a particular location. The nineteenth-century liquor cases tend to suggest that Professor Arcila has the better end of this argument. The Connecticut Supreme Court, in affirming Connecticut's version of the Maine Law, asserted that a bare bones allegation that stolen goods could be located in a particular location satisfied the state constitution's search and seizure provision. Lowrey v. Gridley, 30 Conn. 450, 459–60 (1862).

174. DOCUMENTARY HISTORY, *supra* note 132, at 18.

175. *Id.*

176. *Id.*

Law had described the process in terms that were uniquely related to a liquor search. The search and seizure provision in this version was written in a very generic fashion and resembles language in twentieth-century hornbooks describing the probable cause requirement.

Probable cause in the ordinary criminal justice system was no longer merely a pleading requirement that victims alleged to obtain a warrant; it became the factual threshold that could be satisfied by the testimony of any "credible person." Police officers could satisfy this requirement—and the statute even recognized that police officers could rely on informants to satisfy this requirement.[177] The sworn written statement could offer the facts and circumstances known to the affiant, or the facts and circumstances known to "some other person."[178] In a host of states, the modern probable cause standard for obtaining a search warrant was no longer confined to customs cases, and much of the country embraced the mechanism the courts of Maine developed to remedy and prevent unlawful liquor searches. Failure to comply with the requirements of liquor warrants required exclusion of the fruits of ensuing searches in New York and a number of other states just as it did in Maine.[179]

New York's experiment with Prohibition ended almost as soon as it began.[180] Within a few months of the Maine Law's passage, an Albany jury had acquitted William Landon of violating the Maine Law despite clear

---

177. *See* Davies, *supra* note 61 at 187–88 (probable cause standard made hearsay evidence sufficient for a warrant); *see* Lane, *supra* note 6, at 10–11 (describing the role of early police officers as developing an "intimate familiarity" with the criminals they were policing).

178. DOCUMENTARY HISTORY, *supra* note 132, at 18.

179. *See* People v. Toynbee, 11 How. Pr. 289, 330 (N.Y. Gen. Term 1855) ("The complaint [analogous to the modern affidavit in support of a search warrant] is a substitute for an indictment . . . and requires at least as much particularity . . . ."); *see also* State v. Twenty-Five Packages of Liquor, 38 Vt. 387, 391 (1866) (recognizing that forfeiture action could be quashed for failure to have a sufficiently particular search warrant); Fisher v. McGirr, 1 Gray 1, 2 (Mass. 1854) (action for value of seized liquor permitted on the basis of an insufficient search warrant).

Using an improper search as the basis for dismissing a prosecution would continue into the twentieth century. *See In re* Huff, 120 N.Y.S. 1070 (N.Y. App. Div. 1910) (recognizing that action against forfeited liquor can be dismissed if the search warrant for its discovery is invalid); Foley v. One Hundred & Eighty Bottles of Liquor, 204 N.Y. 623 (N.Y. 1912) (affirming denial of motion to dismiss on this ground). Courts began to expand this rule beyond searches for liquor that were based on invalid search warrants. *See* State v. Kinney, 185 N.Y.S. 645 (N.Y. Sup. 1920) (dismissing indictment for weapon and returning revolver seized by an invalid warrant); People v. Jakira, 193 N.Y.S. 206 (N.Y. Gen. Session. 1922) (gun seized illegally and without warrant excluded).

180. *See* John Joseph Coffey, A Political History of the Temperance Movement in New York State, 1808–1920, at 90–96 (May 1976) (unpublished Ph.D. dissertation, Pennsylvania State) (on file with author); DOCUMENTARY HISTORY, *supra* note 132, at 18; People v. Berberrich, 20 Barb. 168, 266 (1855) (declaring New York's version of the Maine Law unconstitutional).

evidence to the contrary—his lawyer had argued to the jury that the law was unconstitutional.[181] The following year, the New York Court of Appeals agreed.[182] Rather than cure the defects the court identified in the law, the legislature returned to a licensing scheme that strictly regulated who could obtain a license and forbid the sale of alcohol on Sundays.[183] The Maine Law had nevertheless introduced New York's criminal justice system to a search mechanism unmoored from a victim's complaint.

With the new standard came police investigations of victimless crime. Police searches to discover vice in the early days of the New York Municipal Police were rare.[184] The first manual for police officers mentioned the possibility of a search warrant only to discover stolen goods; searches initiated by police to discover evidence of victimless crimes were not mentioned.[185] In the 1870s, private anti-crime organizations began to file complaints seeking arrest and search warrants in cases involving the

---

181. *See* JOHN K. PORTER, ARGUMENT OF JOHN K. PORTER ON THE TRIAL OF WILLIAM LANDON, ACQUITTED JULY 21, 1855, ON A CHARGE OF VIOLATING THE PROHIBITORY LAW (Albany, H.H. Van Dyck 1855) *in* VII AMERICAN STATE TRIALS 901–53 (John D. Lawson, ed., 1917).

182. Wynehamer v. People, 13 N.Y. 378, 486 (1856); *see also* William John Jackson, Prohibition as an Issue in New York State Politics 1836–1933 (February 11, 1974) (unpublished Ph.D. dissertation, Columbia University) (Sacremento State Library).

183. Members of the state legislature were aware that New York Mayor Fernando Wood would thwart efforts to enforce the new licensing regulation, just as he had done with the Maine Law. When they created the new version of liquor violation, legislators replaced the mayor-controlled Municipal Police Force that it had permitted the City of New York to create in 1844 with the Metropolitan Police Force, established under the control of a board appointed by the Governor. An Act of Apr. 15, to Establish a Metropolitan Police District and to Provide for the Government Thereof, New York Laws 1857, ch. 569, 1857 N.Y. Laws 200. The new force was responsible for policing the counties of New York, Kings, Westchester and Richmond, rather than just Manhattan. *Id.*at 200.

184. Prior to the Maine Law, the police had a policy of responding to alleged liquor law violations (i.e., selling without a license or selling on Sunday) only if there was a complaint. *See* RICHARDSON, *supra* note 54, at 110. There was a coordinated raid of brothels in 1850, but other than Mayor Fernando Wood's efforts against lower-class street walkers between 1855 and 1858 (which obviously would not involve the search of any sort of dwelling), there was no substantial subsequent police action against prostitution until the latter part of the century. *See* BURROWS & WALLACE, *supra* note 4, at 807 (describing 1850 raids); *Id.* at 1163 (describing raids of gambling houses and brothels authorized by Mayor Grace in 1886); TIMOTHY J. GILFOYLE, CITY OF EROS: NEW YORK CITY, PROSTITUTION, AND THE COMMERCIALIZATION OF SEX, 1790–1920, at 183–84 (1992); *see also* ANN FABIAN, CARD SHARPS AND BUCKET SHOPS: GAMBLING IN NINETEENTH-CENTURY AMERICA, at 97 (1999) (discussing lack of gambling enforcement from creation of Municipal Police Force through Civil War); RICHARDSON, *supra* note 54, at 154. Given the amount of corrupt coordination between the police and prostitution in the late nineteenth-century, the raids of the late nineteenth century were often more attributable to failure to pay "protection" money to police than the City's serious effort to eliminate prostitution.

185. *See* NEW YORK, RULES AND REGULATIONS, *supra* note 90, at 58.

victimless crimes of pornography, prostitution, gambling and liquor law violations.[186] By the 1880s, a statute specifically authorized police captains to seek warrants to search premises suspected of being houses of prostitution.[187] Police-initiated searches to discover evidence of gambling and alcohol sales without a license, or sales on Sundays occurred with frequency in the latter part of the nineteenth century (most often when the police had not been given their protection money).[188] The word "raid" began to regularly appear in appellate reports by the 1890s.[189]

The Maine Law in New York provided more than just an introduction to a formal search mechanism that could be initiated by someone other than a crime victim. The vigorous debate over the search and seizure aspects of the Maine Law appears to have put to rest any question about the legitimacy of searches initiated by suspicions developed by police officers. Opponents of the Maine Law had objected both to the prohibition of alcohol sales and the expansion of police discretion. In the latter half of the nineteenth century, there were proposals to regulate rather than prohibit

---

186. *See* Dan Greenberg & Thomas H. Tobiason, *The New Legal Puritanism of Catherine MacKinnon*, 54 OHIO ST. L.J. 1375, 1377–78 (1993) (citing Felice F. Lewis, Literature, Obscenity, & Law 10 (1976)); Louis H. Pollack, Review of: Federal Censorship: Obscenity in the Mail, 75 HARV. L. REV. 1681 (1962).

187. People *ex rel.* Eakins v. Roosevelt, 44 N.Y.S. 1003 (N.Y. 1897). This was one example of many in the late nineteenth century of members of anti-vice societies who went undercover to discover prostitution and its lack of enforcement. *See e.g.*, GILFOYLE, *supra* note 184, at 181–96 (describing these societies). The Court's opinion in *Eakins* colorfully describes, using appropriately prudish language, the adventures of the undercover member of an anti-vice society who discovered a brothel and reported the failure of police to close it. The citizen-informant entered a dwelling between 2:00 and 3:00 am one morning, saw 16 to 20 women huddled around a few men while women continued to enter and leave the room and went upstairs with one of the women upon his payment of the 25 cents rent for the room. On his way to the room, he "saw and heard the most disgusting evidence of vice." *Eakins*, 44 N.Y.S. at 1007.

Technically, prostitution itself was not a crime in New York in the second half of the nineteenth century, only the crime of maintaining a house of prostitution. *See* THOMAS C. MACKEY, RED LIGHTS OUT: A LEGAL HISTORY OF PROHIBITION, DISORDERLY HOUSES AND VICE DISTRICTS, 93–118 (1987); William J. Stuntz, *The Pathological Politics of Criminal Law*, 100 MICH. L. REV. 505, 573–74 (2001) ("Before the late nineteenth century, most jurisdictions had no prostitution statutes; the relevant crime was running a 'disorderly house,' a more circumscribed offense.").

188. Police corruption was rampant in the late nineteenth century. Bars, gambling houses, and brothels regularly paid police to avoid prosecution. EDMUND MORRIS, THE RISE OF THEODORE ROOSEVELT 499 (2001).

189. *See* Mott v. Mott, 38 N.Y.S. 261, 262 (N.Y. 1896) (divorce action in which husband's alleged adultery with a prostitute was testified to by a woman who saw the man in an apartment she kept; the husband was in the apartment a day or two before a police raid); People ex rel. Doherty v. Police Com'rs of New York, 84 Hun. 64, 66 (1895) (operator of house of ill fame claimed that officer extorted money from her so that she could avoid being raided again); People ex rel.. Cross v. Martin, 32 N.Y.S. 933 (N.Y. 1895).

gambling and prostitution.[190] There were advocates of legalized gambling and prostitution, just as there were for lawful alcohol sales.[191] There were, however, no objections to the power of the police to conduct the raids necessary to enforce the prohibition on these vices.[192] Police-initiated and police-conducted investigations had come to be accepted by the latter part of the nineteenth century. Probable cause that could be satisfied by any person with relevant information, including officers, had become an unquestionably sufficient criterion for police searches.

III. LAW ENFORCEMENT NEEDS PROMPTED A BROADER ARREST POWER

The police had no obvious ally in advocating a less restrictive standard for arrests. While prohibitionists and police wanted relaxed standards for conducting searches—prohibitionists to discover liquor, police to discover evidence of crime more generally—no analogous group shared the interest of police in readily being able to take suspects into custody. The success of the more police-friendly arrest standard turned alone on policymakers' interest in giving police greater discretion and their ability to exercise that discretion responsibly. The emergence of probable cause as a standard sufficient for arrest occurred more slowly in New York than in other

190. A variety of laws were passed in the nineteenth century that allowed gambling in certain circumstances. *See* BURROWS & WALLACE, *supra* note 4, at 1164. The idea of legalizing prostitution received less serious attention but was nevertheless considered. *See* RICHARDSON, *supra* note 54, at 154 (discussing post-Civil War discussion to license and regulate prostitution as Union army had done in the occupied City of Nashville, Tennessee). For discussion of the nation's first legalized prostitution, *see* THOMAS PETER LOWRY, THE STORIES THE SOLDIERS WOULDN'T TELL: SEX IN THE CIVIL WAR 76–82 (1994); WALTER T. DURHAM, RELUCTANT PARTNERS: NASHVILLE AND THE UNION, JULY 1, 1863 TO JUNE 30, 1865 (1987). Contemporaneous with General Hood's order in Nashville, Great Britain's Contagious Disease Act of 1864 effectively legalized prostitution. *See* BURROWS & WALLACE, *supra* note 4, at 1162.

191. Prostitution and gambling appealed to a smaller audience—at least a smaller audience willing to publicly associate themselves with these acts—than drinking. *See* Michael Woodiwiss & Dick Hobbs, *Organized Evil and the Atlantic Alliance: Moral Panics and the Rhetoric of Organized Crime Policing in American and Britain* 49 BRIT. J. CRIMINOLOGY 106, 109 (2009) (observing that proponent of legalized gambling lacked the financial wherewithal of the advocates of repealing Prohibition).

192. The procedural requirements for warrants that the Maine Law had introduced remained. In 1891, an appellate division of the New York Supreme Court held that an affidavit for a search warrant failed to state the facts supporting the affiant's suspicions that his stolen goods could be found in the location to be searched. The court further held that this failure in the affidavit prevented the justice of the peace from obtaining jurisdiction to issue the warrant, leaving him liable to a civil action. Wallace v. Williams, 14 N.Y.S. 180 (N.Y. Gen. Term 1891). Remarkably, this case involved a warrant to locate stolen goods, the paradigm search warrant in the eighteenth century, which had been authorized on a victim's mere assertion that he had probable cause to believe (or, before the adoption of revised statutes in several states, suspect) that the goods could be located in the place identified.

jurisdictions, and a quirk of legislative timing reversed the broader standard previously conferred on officers. Nevertheless, New York's experience is representative in illustrating the forces at play in expanding the arrest powers of police and how tenuously police departments held these powers in their early years.

The fear New Yorkers had of police power certainly did not dissipate with the creation of the Municipal Police Department in 1845.[193] Though the new department advocated legal reform to permit officers to arrest on mere probable cause of a felony, neither the courts of New York, nor the New York Legislature, were initially willing to embrace this new arrest standard. In 1853, New York courts began to accept probable cause as a sufficient basis for a warrantless felony arrest.[194] There was, however, an important difference between the probable cause standard for search and arrest warrants and the probable cause standard for warrantless arrests. Officers, not magistrates, obviously determined whether probable cause existed to justify a warrantless arrest, and these officers were known for rampant violence and arbitrary arrests when the New York Legislature adopted its Code of Criminal Procedure in 1881.[195] By considering this code in the early 1880s, New York politicians were forced to take a stand against the police during a period of fairly serious misconduct. By the 1890s, Progressive reformers had successfully blamed police misconduct on a culture of corruption and proposed good government reforms as a cure.[196] In the early 1880s, however, a limit on the discretion of officers to make warrantless arrests may well have seemed to be a decent remedy for arbitrary arrests.[197]

Americans were generally less willing, or at least slower, than their English counterparts to expand the discretionary powers of police officers—New Yorkers would appear especially unwilling to extend prerogatives to them. The King's Bench adopted the rule American lawyers would presently recognize as the standard for warrantless arrests in 1827—if an officer had probable cause to believe that a crime had been committed and that the person taken into custody committed it, the officer is not liable for false arrest even if the suspect he took into custody was factually innocent.[198] This modification of the English law may have been prompted by a perception that constables needed more tools at their disposal to deal

193. New York's force represented the second full-time patrol force, second only to London's modern police force, created in 1829. WILBUR R. MILLER, POLICE AUTHORITY IN LONDON AND NEW YORK CITY 1830–1870, 8 JOURNAL OF SOCIAL HISTORY 81, 81 (1975).

194. Burns v. Erben, 40 N.Y. 463 (N.Y. 1869).

195. *See* JOHNSON, *supra* note 4, at 15–16.

196. *See* Oliver, *supra* note 57, at 468–83. *See generally* Jay S. Berman, POLICE ADMINISTRATION AND PROGRESSIVE REFORM: THEODORE ROOSEVELT AS POLICE COMMISSIONER OF NEW YORK (1987).

197. JOHNSON, *supra* note 4 at 15–16.

198. Beckwith v. Philby, (1827) 108 Eng. Rep. 585 (K.B.); 6 Barn. & Cress. 635.

with increasing concerns of crime and violence, particularly in urban areas. The new arrest rule was not, however, the product of pressure from, what we would call in modern times, the law enforcement lobby. The Metropolitan London Police Department would not be established for another two years.[199]

A number of American states followed the English precedent, adopting the new standard before the creation of metropolitan police forces. Cases in Massachusetts and Pennsylvania adopting probable cause as the standard for a warrantless arrest pre-dated the creation of modern police departments in Boston and Philadelphia by a few years.[200] Tennessee adopted the probable cause standard decades before the creation of modern police forces in Nashville or Memphis.[201] New York did not, however, accept this standard as a basis for warrantless arrests until almost a decade after the creation of the Municipal Police Force.

The New York Muncipal Police Department was created in 1846 and its officers were instructed that they could arrest any "person who has committed a felony, or who for reasonable cause, is suspected of having committed a felony."[202] As late as 1852, however, Oliver Barbour's treatise on New York criminal law observed that officers were permitted to arrest on the basis of probable cause only if a felony had in fact been committed.[203] It was not until 1853 that two justices of the New York Supreme Court, the trial level court, in this case sitting in New York City, acknowledged in dicta the power of officers to arrest a felony suspect when there is "just suspicion."[204] While this probable cause arrest standard would be accepted in New York, at least for a few decades, it was certainly not an uncontroversial standard.

Burns v. Erben, decided by a three-judge appellate panel of the New York Superior Court in 1864, would reveal that New York courts were not entirely comfortable with the new arrest standard.[205] The court did not have

---

199. Craig D. Uchida, THE DEVELOPMENT OF THE AMERICAN POLICE: AN HISTORICAL OVERVIEW, 7 (2004).

200. Rohan v. Swain, 59 Mass. (5 Cush.) 281, 282 (1850); Russell v. Shuster, 8 Watts & Serg. 308, 309 (Pa. 1844); Eanes v. State, 25 Tenn. 53, 54 (Tenn. 1845).

201. *See Eanes*, 25 Tenn. at 54.

202. NEW YORK, RULES AND REGULATIONS, *supra* note 90, at 31.

203. OLIVER L. BARBOUR, A TREATISE ON THE CRIMINAL LAW OF THE STATE OF NEW YORK (Albany, Gould, Banks 2d ed. 1852).

204. Pratt v. Hill, 16 Barb. 303 (N.Y. Sup. Ct. 1853). *Pratt* cites *Holley v. Mix*, 3 Wend. 350, 353 (N.Y. 1829) for the proposition that an officer's suspicion was sufficient for a felony arrest, but as Thomas Davies observes, *Holley* was a case recognizing only an officer's immunity for arresting without a warrant when a citizen charged that the suspect had committed a crime. Davies, *supra* note 5, at 635 n.239.

205. There was a passing reference in dicta to officers being permitted to arrest on mere probable cause the year before *Burns* was decided. *See* Brown v. Chadsey, 39 Barb. 252, 263 (NY Supreme Ct, NY County 1863) ("probable cause, or reasonable grounds of suspicion against the party arrested, afford no justification of an arrest or imprisonment which is

to acknowledge the legitimacy of the probable cause standard to rule in favor of the officer in this case in which the plaintiff alleged a wrongful arrest. The officer in *Burns* made the arrest after a complainant alleged the suspect had stolen his property.[206] The officer was justified under well-established law, as a crime had in fact been committed. The officer was, however, represented by the Corporation Counsel for the City of New York, who hoped to use this case to clearly establish the new arrest standard. The City's lawyer observed that the law creating the New York Metropolitan Police had given members of the new police force the warrantless arrest powers of constables, which he argued included the power to arrest when an officer had probable cause.[207] In support of his description of a constable's authority, counsel offered the authority of *Beckwith v. Philby* from 1827, which several American jurisdictions had adopted but New York courts had not embraced.[208]

The court accepted the probable cause arrest standard the City advocated, but its reasoning differed from the City's in an important respect. The court observed that the "Metropolitan Police Act allow[ed] the officers of police to arrest persons suspected by them, without warrant, where there is reason to believe a felony has been committed."[209] The court therefore attributed the probable cause standard to the statute itself, concluding that the statute itself embraced the probable cause standard. The statute of course only gave officers the power of constables, but the court appears to have been reluctant to interpret the power of constables to include this standard. If this (elected) court accepted the English precedent, however, it would bear the responsibility for defining the power of constables, something it likely did not want to appear to do. It was far more comfortable attributing this standard, which it recognized to entrust a "dangerous power" in the police, to the legislature.[210]

There were certainly strong supporters of the probable cause arrest standard in New York who were less sheepish. A three-judge panel of the New York Supreme Court for New York County was bolder in its reasoning one year later when it affirmed the rule announced in *Burns*. The court questioned rhetorically, "How, in the great cities of the land, could police power be exercised, if every police officer is liable to a civil action for false imprisonment, if persons arrested upon probable cause shall afterwards be found innocent? Police authority would be a sham, its officers be made cowards, and government become a failure."[211]

---

without authority of law.").

206. Burns v. Erben, 26 How. Pr. 273 (N.Y. Super. Ct. 1864).

207. *Id.*

208. *See* Davies, *supra* note 5, at 636 (discussing Holley v. Mix, 3 Wend 350 (1829)).

209. *Burns*, 26 How. Pr. 273.

210. *Id.*

211. Hawley v. Butler, 54 Barb. 490, 496 (N.Y. Sup. 1868).

The fear of officers' discretion was, however, particularly strong in the second half of the nineteenth century in New York. When the *Burns* case was appealed to the New York Court of Appeals, it was clear that serious questions remained about conferring discretion on officers to evaluate whether evidence was sufficient to justify a warrantless arrest. Two judges wrote opinions in *Burns,* each finding the arrest acceptable, though neither would accept the probable cause arrest standard. Judge Woodruff concluded that a warrantless arrest by an officer was acceptable, even if no crime had been committed, if the officer "acted upon information from another which he had reason to believe."[212] In the early nineteenth century, an officer's powers had been expanded to permit a warrantless arrest if a complainant had made a positive charge against the would-be arrestee—in other words, something analogous to a victim's complaint to a magistrate.[213] Under this long-standing justification for a warrantless arrest, not the new probable cause standard, Woodruff found the officer's actions justified. Judge James was more clear in his refusal to endorse the new arrest standard. He also wrote an opinion in *Burns*, concluding that "[p]robable cause, or reasonable ground, for suspicion . . . affords no justification for an arrest or imprisonment, unless a felony has actually been committed."[214]

New York's particular concern with police powers is difficult to explain.[215] New York's unique history may offer some insight. Historians typically explain early concerns about the powers of modern American police as a manifestation of Revolutionary-Era fears of standing armies.[216] Lingering fears about standing armies seem to have had particular salience in New York—the concern had successfully thwarted the effort to create a London-style modern police force in the 1830s.[217] Certainly there was a fear of standing armies throughout the young republic, but there may be a reason that the analogy to modern police forces got particular traction in New York.[218] Only New York, Philadelphia, and Boston had experienced British occupation during the Revolutionary War. Boston had been occupied for only eleven months,[219] Philadelphia for nine,[220] while New

212. Burns v. Erben, 40 N.Y. 463, 469 (N.Y. 1869).

213. *See* Davies, *supra* note 5 at 650–54.

214. *Burns*, 40 N.Y. at 466.

215. While New York's resistance to the probable cause arrest standard is unique among states with large urban populations in the mid-nineteenth century, New York was certainly not alone in having concerns about the new standard. *See* Davies, *supra* note 5, at 637 and n.246 (describing a North Carolina Supreme Court Justice's resistance to the probable cause arrest standard).

216. *See* BURROWS & WALLACE, *supra* note 4, at 636; RICHARDSON, *supra* note 54, at 25.

217. BURROWS & WALLACE, *supra* note 4, at 636.

218. *See* RICHARDSON, *supra* note 54, at 15.

219. DAVID MCCULLOUGH, 1776 25 (2005) (describing the siege and occupation of Boston).

220. JOHN W. JACKSON, WITH THE BRITISH ARMY IN PHILADELPHIA, 1777–1778 351

York was occupied for seven years during which time the houses of New Yorkers were frequently plundered.[221]

Legislators were required to weigh in on the probable cause standard in 1881 for reasons that appear to be accidental, or at least entirely unrelated to anything related to warrantless arrests. When New York codified its laws in 1829, the legislature provided for a Commission on Pleading and Practice to draft a Code of Civil Procedure and a Code of Criminal Procedure.[222] The Commissioners' proposal was considered in 1849, 1850, and 1855, but never adopted.[223] The legislature considered the Code again in 1881. The Code delineated rules for all aspects of the criminal justice system, not just the police. It did, however, include a variety of rules regarding officers, including the arrest standard.[224]

From the perspective of the police, it was particularly bad timing for elected officials to be publicly considering rules involving their discretion. Reports of police brutality became frequent in the late 1860s and continued to escalate into the 1870s.[225] As one might expect, working class New Yorkers were most frequently the targets of acts of official violence. While upper class New Yorkers tended to appreciate the peacekeeping role of police, working class New Yorkers tended to have some degree of fear of the new institution.[226] The class tensions in policing were aggravated by the police department's violent relationship with organized labor. Police efforts to contain labor demonstrations in the latter half of the nineteenth century frequently resulted in violence. Clubs were often used to break up strikes and protests.[227]

One such confrontation left long memories. In 1874, several labor organizations planned a rally in Tompkins Square.[228] Permits were required in New York after 1872 for any sort of public meeting.[229] The groups were initially granted permission to hold their event, but then the permits were revoked the night before the event because of the concern police had that "the proposed meeting would endanger the public peace."[230] The concerns

---

(1979) (describing occupation of Philadelphia).

221. JUDITH L. VAN BUSKIRK, GENEROUS ENEMIES: PATRIOTS AND LOYALISTS IN REVOLUTIONARY NEW YORK 23 (2002) (describing occupation of New York).

222. *See* John T. Fitzpatrick, *Proposed Codes of the State of New York*, LAW LIBR. J., 12, 20 (1924).

223. *See id.*

224. *See* CRIMINAL CODE, *supra* note 37, at 88–89.

225. JOHNSON, *supra* note 4, at 17–18.

226. *Id.* at 30–38.

227. *Id.* at 30.

228. At the time of the demonstration, New York City was overpopulated with out-of-work and homeless individuals. *See* Luc Sante, LOW LIFE: LURES AND SNARES OF OLD NEW YORK 354 (2003). An estimated 110,000 out-of-work individuals and 10,000 homeless resided in the city. *See id.*

229. KELLER, *supra* note 90, at 174.

230. *Id.*

of the police were not utterly unfounded. Various groups that were to participate in the event had accused one another of having dangerous affiliations, fostering fears that labor groups were inciting revolution.[231] The last-minute revocation of the license was not, however, a recipe for minimizing civil unrest. Many participants arrived for the event unaware that the permit had been revoked. Without telling the crowd to disperse, officers rushed into the crowd of 1,500 demonstrators with horses and clubs, battering an untold number with locust clubs and arresting forty-four on charges varying from disorderly conduct and incendiary speech to assault and battery.[232] An editorial in the *New York Herald* stated, "the average policeman, running a muck [*sic*] with his locust in hand, is not to be relied on for the exercise of much discretion."[233]

There were certainly defenders of the police after the Tompkins Square Riot who applauded the maintenance of order, just as police supporters had always done, but there was a growing sense that the police were out of control.[234] Newspapers increasingly reported random acts of violence by police.[235] A number of seemingly innocent citizens were clubbed while sitting on their front stoops in the mid-1870s, leading the *New York Times* to describe "The Front Steps Crime."[236] No clear consensus emerged on how to deal with the problem. Working class New Yorkers called for stricter regulation of the police while middle and upper class New Yorkers regarded police violence as the symptom of a larger problem of official corruption.[237] Working class New Yorkers also complained of corrupt

---

231. *Id.* at 173.

232. *Id.* at 174. Tompkins Square is no stranger to dramatic events. Ironically, there were two other riots in Tompkins Square in 1988 and 1995, aptly labeled "Tompkins Square Park Riot II" and "Tompkins Square Park Riot III." BRIAN ST. CLAIRE-KING, FATES WORSE THAN DEATH 405 (2003). Riot II resulted when police attempted to evict homeless individuals from the park. *Id.* When protesters showed up, several police placed tape over their badge numbers and began beating them up. *Id.* Riot III also occurred when police attempted to evict the homeless. *Id.*

233. KELLER, *supra* note 90, at 175. News quickly spread of the Tompkins Square Riot. Only an hour after the rioting began, the *New York Graphic* published a headline reading: "A Riot Is Now In Progress in Tompkins Square." MICHAEL SORKIN, VARIATIONS ON A THEME PARK 67 (1992).

234. KELLER, *supra* note 90, at 174–75; JOHNSON, *supra* note 4, at 32. Police blamed the riot on "Parisian 'Communists,' 'heavily armed German revolutionaries,' 'atheists,' and 'drunkards.'" M. J. HEALE, AMERICAN ANTICOMMUNISM: COMBATING THE ENEMY WITHIN, 1830–1970, 25 (1990). Following the riot, New York City police sent special detectives into socialist and labor meetings as spies. FRANK DONNER, PROTECTORS OF PRIVILEGE: RED SQUADS AND POLICE REPRESSION IN URBAN AMERICA 11 (1992). Meanwhile, the New York police board alleged that radicals were planning on burning down churches and accumulating firearms and ammunition for "a bloody showdown." *Id.*

235. JOHNSON, *supra* note 4, at 18, 38–39.

236. *Id.* at 39.

237. *Id.* at 39–41. New York City was no stranger to public rioting. The city endured

officers profiting from extortion while brutalizing citizens, but rather than criminal sanctions, they understandably wanted the more direct remedy of limits on discretion.[238]

By the 1890s, progressive reformers began to express more compassion toward the working-class victims of police violence and won public support for their view that good government, anti-corruption measures held the cure to police violence.[239] The Republican-led Wickersham Commission in 1895 provided a supportive forum for working class citizens to publicly describe the abuse they suffered at the hands of police.[240] In the 1880s, however, there were conflicting views of the appropriate remedy to very substantial problems within the police department. Increasing the discretion of officers to decide when to arrest, in light of patterns of abuse, was not politically expedient at a time when a number of New Yorkers were calling for greater restraints.

Even though the proposed code the legislature considered in 1881 was first drafted in 1849, when the police department was in its infancy, the final version the legislature adopted was far less favorable to the police than the original draft. The 1849 version of the code included all of the justifications for a warrantless arrest that had been recognized in the United States before the creation of professional police forces. Under the proposed code, an officer could arrest a suspect without a warrant:

1. For a public offense, committed or attempted in his presence;
2. When the person arrested has committed a felony, although not in his presence;
3. When a felony has in fact been committed, and he has reasonable cause for believing the person arrested to have committed it; [and]
4. On a charge, made upon reasonable cause, of the commission of a felony by the party arrested.[241]

---

riots in 1806, 1826, 1834, 1837, 1849, 1855, 1857, 1863, 1870, 1874, and 1900. Eric H. Monkkonen, *History of Urban Police*, 15 CRIME AND JUSTICE 547, 553 (1992).

238. JOHNSON, *supra* note 4, at 33.

239. The New York legislature created the Lexow Commission in 1894 to investigate New York police. Monkkonen, *supra* note 237, at 565 (citing JAY S. BERMAN, POLICE ADMINISTRATION AND PROGRESSIVE REFORM: THEODORE ROOSEVELT AS POLICE COMMISSIONER OF NEW YORK 23–29 (1987)). Lexow was not the last commission to work against police corruption. It was followed by the Curran Committee of 1913, the Seabury Investigation of 1932, and the Hefland Investigation of 1955. Dean Joan Wexler, *Police Violence: Causes and Cures*, 7 J.L. & POL'Y 75, 77 n.1 (1998).

240. JOHNSON, *supra* note 4, at 133–41.

241. CRIMINAL CODE, *supra* note 37, at 88–89. The language in this proposed statute is awkward as a prescription for officers, as it was taken from decisions involving suits for unlawful arrests. Exceptions (2) and (3) provide the officer immunity from civil liability if the suspect is in fact guilty, or if a crime was in fact committed, something that could not be known with certainty. Practically speaking then, an officer would be willing to arrest only if

The Commissioners in 1849 additionally proposed permitting officers to make an arrest for a felony at night, even if it should "afterwards appear that a felony had not been committed."[242] The Commissioners appear to have embraced the ages-old perceived need to allow greater security at night.[243] Under English, colonial and early American laws, persons who could not explain their presence on the streets of a town at night could be detained until they were taken before magistrates to explain themselves.[244]

The Code of Criminal Procedure adopted this warrantless standard from the proposed code but eliminated the fourth exception that allowed an officer to arrest on a complainant's charge.[245] The New York Legislature had restored the very restrictive arrest standards that governed constables and watchmen in the eighteenth century. There is certainly an irony to this. New York had the largest police force in the country in the 1880s, and no other legislature had moved to restrict the discretion of police.[246] Just as New York's particular history explained the reluctance of its judiciary to adopt the probable cause standard for warrantless arrests, events occurring only in New York set the state's police regulation apart. Police violence was certainly not confined to New York in the latter part of the nineteenth century, but no other state with a modern metropolitan police force drafted a criminal procedure code in the 1880s.[247] Unlike other state legislatures, New York's was forced to take the public's pulse on police regulation in 1881, while the legislatures of other states could sit on the sidelines as courts continued to rely on mid-nineteenth century precedent.

The unease with broad arrest powers New Yorkers demonstrated in 1881 likely was not limited to residents of the Empire State, but the timing

---

the facts giving rise to his suspicion occurred before his eyes, as described in exception (1), or a victim made a complaint to him, as in exception (4), or he had some extraordinary basis for suspecting the arrestee. *See generally* Davies, *supra* note 5.

242. CRIMINAL CODE, *supra* note 37, at 89; *see* Atwater v. Lago Vista, 532 U.S. 318, 333 (2001) (quoting 2 Hawkins, ch. 13, §6, at 130) ("'[I]n affirmance of the common law,' for 'every private person may be the common law arrest any suspicious night-walker, and detain him till he give good account of himself.'").

243. The nightwalker statutes are not the only ones that recognize additional security concerns at night. Burglary at night has long been recognized as a more serious crime than burglary during the day. *See* Theodore E. Lauer, *Burglary in Wyoming*, 32 LAND & WATER L. REV. 721, 726–29 (1997) (describing history of crime of burglary in England, noting distinction between entering a dwelling at night and day).

244. *See Atwater*, 532 U.S. at 333 (citing to sources recognizing nightwalker statutes).

245. N.Y. CODE CRIM. PROC. Law. § 177 at 52 (Gould 1881). The legislature did, however, retain the right of an officer to arrest at night on mere probable cause. *See* N.Y. CODE CRIM. PROC. Law § 179 at 52 (Gould 1881).

246. Johnson, *supra* note 4, at 2.

247. This accident of history seems to explain why New York was alone in the mid-twentieth century in requiring that a crime have been committed in fact before an officer could make a felony arrest on probable cause alone. *See* Davies, *supra* note 5, at 578 (noting the anomaly).

of the Code of Criminal Procedure uniquely memorialized the public's view at a point when police power was particularly feared. The new limitation, however, appears to have had only a minimal effect on the police department, as relatively few cases can be located in which officers were sued for arresting a suspect when no felony had in fact been committed.[248] All the while, the more police-friendly probable cause arrest standard was gaining acceptance outside New York despite the concerns about arbitrary arrests and police brutality raised by the creation of modern police forces.[249] The need for greater police authority—to control the streets and investigate crimes—had ushered in a new arrest standard.

CONCLUSION

Probable cause, as we understand it today, was not a sufficient basis for a law enforcement officer to make an arrest or seek a search warrant in late eighteenth-century America. However probable cause, as we understand it today, was more than sufficient for a *victim* to seek a search or arrest warrant, or instruct an officer to make an arrest. Probable cause was, in essence, a pleading requirement for victims. Law enforcement officers, by contrast, were required to observe the crime in progress, or wait for a victim's complaint, before they could even seek a magistrate's authorization for a search or arrest.

The modern understanding of probable cause is an evidentiary threshold that may be satisfied by any person with information "sufficient to warrant a person of reasonable caution in believing that a crime ha[s] been committed" or is about to be committed.[250] This evidentiary threshold may be satisfied by any person with evidence bearing on the question of whether there is suspicion; the modern standard does not depend on the identity of the person claiming to have probable cause or the type of crime investigated. More is required of victims than was required during the Framing Era and less is required of law enforcement than was required during the Framing Era. Victims must demonstrate the basis of their suspicion while law enforcement officers are no longer dependent on victims.

The need for greater security forced society to trust law enforcement officers with greater discretion—a trust that was not readily granted and not well-earned. At the same time the realities of urban life were forcing Americans to place the same faith in law enforcement officers that they placed in private citizens, the public began to lose its faith in the integrity of

---

248. The first appellate case on this issue following the 1881 Code appears to have been *Stearns v. Titus*, 85 N.E. 1077 (1908). Carolyn Ramsey has noted a similar ineffectiveness of tort suits to constrain the practice of material witness detentions in the nineteenth century as potential litigants do not appear to have brought actions. Ramsey, *supra* note 5, at 703–04.

249. *See* Wilgus, *supra* note 56, at 818–20.

250. New Jersey v. T.L.O., 469 U.S. 325, 366 n.7 (1985).

private complainants. The citizenry was unwilling to entrust Temperance Watchmen, the teetotaling private citizens who sought warrants against their less rigid neighbors, to direct the state's searching apparatus. The citizen-informants were disliked for their zeal that led them to make allegations on less than reliable evidence as much as they were for their thorough investigations accurately identifying liquor law violators. Their enthusiasm for a despised law thus prompted a new, more heavily scrutinized method for citizen-requested searches.

Probable cause is, of course, something of a universal standard for authorizing searches or arrests in the twenty-first century and has been for some time. But the standard's ubiquitous quality is of more recent origin than a reading of the Supreme Court's criminal decisions—or even the text of the Constitution itself—might suggest.

WHY CIVIL RECOURSE THEORY IS INCOMPLETE

CHRISTOPHER J. ROBINETTE*

I. INTRODUCTION 432
II. CIVIL RECOURSE EXPLAINED 433
A. The Basic Structure 433
1. Wrongs 435
a. Relational 435
b. Guidance Rules 436
c. The Source of the Wrongs 436
2. Rights of Redress 439
a. Satisfaction 439
b. Preservation of Civil Society 440
B. The Historical Account 440
III. INSTRUMENTALISM IN TORT LAW 444
A. The Concessions 445
B. History Revisited 447
1. The Collapse of the Writ System 447
2. Disentangling Contract 448
3. The Creation of Negligence as an Independent Cause of Action 449
4. The Surge in Accidental Injuries During the Industrial Revolution 450
5. The Advent of Liability Insurance 452
6. An Ethic of Interdependence 453
C. Doctrine Revisited: Tort Reform 456
1. Workers' Compensation 458
2. Automobile Accidents 463
3. Products Liability 471
4. "Modern" Tort Reforms 475
IV. CONCLUSION 484

* Associate Professor, Widener University School of Law. B.A. 1993, College of William & Mary; J.D. 1996, University of Virginia School of Law. I thank Ken Abraham, Ben Barros, Alan Calnan, Jonathan Cardi, Mike Dimino, Tom Grey, Juliet Moringiello, Jeffrey O'Connell, Wes Oliver, Mike Rustad, Sheila Scheuerman, Victor Schwartz, Jane Stapleton, Byron Stier, and John Witt for reviewing a draft of this article or assistance with specific issues. I also benefited from participating on a panel at the 13th Annual Conference of the Association for the Study of Law, Culture, and the Humanities at Brown University. I am grateful to Widener University for generous support of this Article. For excellent research assistance, I thank Stefanie Pitcavage, Kirsten Kutler, and Ulysses Wilson.

I. INTRODUCTION

What is tort for? From the days of Oliver Wendell Holmes, Jr. to the present, many scholars have pointed to a single answer.[1] Holmes began his chapters on torts in *The Common Law* by stating: "The object of the next two Lectures is to discover whether there is any common ground at the bottom of all liability in tort, and if so, what that ground is."[2] The latest attempt to so unify tort law is the rich "civil recourse" theory of Professors John C.P. Goldberg and Benjamin C. Zipursky.[3]

Civil recourse has both a positive and a negative claim.[4] The positive claim is that tort law is about wrongs.[5] Goldberg succinctly argues that tort's purpose is "providing victims with an avenue of civil recourse against those who have wrongfully injured them."[6] The negative claim is that tort has no other purposes.[7] Many other scholars describe tort law as a means to compensate injured victims and spread losses[8] or deter accidents with the threat of liability.[9] By contrast, Goldberg and Zipursky deny that tort is designed to serve these—or any—independent social or public policy goals.[10] Tort is not a mere instrument: "[W]e must recapture the idea that

---

1. *See, e.g.*, Christopher J. Robinette, *Can There Be a Unified Theory of Torts? A Pluralist Suggestion from History and Doctrine*, 43 BRANDEIS L.J. 369, 369–70 (2005); Christopher J. Robinette, *Torts Rationales, Pluralism, and Isaiah Berlin*, 14 GEO. MASON L. REV. 329, 329 (2007).

2. OLIVER WENDELL HOLMES, THE COMMON LAW 63 (Mark DeWolfe Howe ed., Harvard Univ. Press 1963) (1881).

3. *See* John C.P. Goldberg & Benjamin C. Zipursky, *Accidents of the Great Society*, 64 MD. L. REV. 364 (2005) [hereinafter Goldberg & Zipursky, *Accidents of the Great Society*]; John C.P. Goldberg & Benjamin C. Zipursky, *The Moral of MacPherson*, 146 U. PA. L. REV. 1733 (1998) [hereinafter Goldberg & Zipurksy, *Moral of MacPherson*]; John C.P. Goldberg & Benjamin C. Zipursky, *Tort Law and Moral Luck*, 92 CORNELL L. REV. 1123 (2007) [hereinafter Goldberg & Zipursky, *Moral Luck*]; John C.P. Goldberg & Benjamin C. Zipursky, *Torts as Wrongs*, 88 TEX. L. REV. 917 (2010) [hereinafter Goldberg & Zipursky, *Torts as Wrongs*]; John C.P. Goldberg & Benjamin C. Zipurksy, *Unrealized Torts*, 88 VA. L. REV. 1625 (2002) [hereinafter Goldberg & Zipursky, *Unrealized Torts*]; Benjamin C. Zipursky, *Civil Recourse, Not Corrective Justice*, 91 GEO. L.J. 695 (2003) [hereinafter Zipursky, *Civil Recourse, Not Corrective Justice*]; Benjamin C. Zipursky, *Rights, Wrongs, and Recourse in the Law of Torts*, 51 VAND. L. REV. 1 (1998) [hereinafter Zipursky, *Rights, Wrongs*].

4. *See* discussion *infra* Parts II.A, III.

5. *See infra* notes 14–17 and accompanying text.

6. John C.P. Goldberg, *Ten Half-Truths About Tort Law*, 42 VAL. U. L. REV. 1221, 1252 (2008).

7. *See infra* notes 99–100 and accompanying text.

8. *See, e.g.*, DON DEWEES ET AL., EXPLORING THE DOMAIN OF ACCIDENT LAW 5–8 (1996); MARK C. RAHDERT, COVERING ACCIDENT COSTS 101, 173–74 (1995).

9. *See, e.g.*, RICHARD POSNER, ECONOMIC ANALYSIS OF LAW 143 (2d ed. 1977); George Priest, *Modern Tort Law and Its Reform*, 22 VAL. U. L. REV. 1, 20–21 (1987).

10. John C.P. Goldberg, *Unloved: Tort in the Modern Legal Academy*, 55 VAND. L.

tort cases are concerned with the focused task of identifying and remedying instances in which an actor has wronged another, as opposed to providing localized compensation or insurance schemes, regulating antisocial conduct for the good of society, or the like."[11]

Although I agree with Goldberg and Zipursky's positive claim that tort is about wrongs, their exclusion of instrumentalist factors such as compensation, deterrence, and administrative efficiency is erroneous. These other elements are also important to tort law; therefore, civil recourse theory alone provides an incomplete explanation for it. Surprisingly, the first evidence of this theory's shortcomings comes from Goldberg and Zipursky themselves.[12] In several footnotes scattered throughout different articles, Goldberg and Zipursky have conceded that aspects of multiple tort doctrines cannot or may not be conceived of as wrongs.[13] These concessions beg the question of whether even more instrumentalism is operating in tort law. I pursue this issue through the lens of tort reforms over the last century.

In that vein, I describe civil recourse theory and its account of wrongfulness in Section II. In Section III, I enumerate those doctrines that Goldberg and Zipursky admit do not fall within their theory. The instrumentalist rationales of deterrence or compensation explain these doctrines. I then examine tort reform over the last century. The course of tort reform demonstrates the development of instrumentalism within tort law, either in the rationales of compensation and deterrence or in the instrumentalist concern of administrative efficiency. All of the major tort reforms over the last century—workers' compensation, no-fault automobile reform, products liability, and "modern" reforms—were based in instrumentalism. Moreover, when the reforms are viewed chronologically, a pattern develops: In each successive reform, instrumentalism made increasing inroads into tort. Civil recourse theory, in failing to acknowledge this instrumentalism, omits a substantial component of tort law. Finally, in Section IV, I conclude with a discussion of tort from a normative perspective.

II. CIVIL RECOURSE EXPLAINED

A. The Basic Structure

Civil recourse is a positive, as opposed to a normative, theory of tort law;[14] it purports to describe tort as it is, not as it should be.[15] The theory

---

REV. 1501, 1519 (2002).

11. *Id.*

12. *See infra* notes 101–10 and accompanying text.

13. *See infra* notes 101–10 and accompanying text.

14. *See, e.g.*, Robert W. Bennett, *Response: On Substantiation of Positive Social Theory*, 95 NW. U. L. REV. 977, 983 (2001) ("A normative account provides an ideal to be

has two primary components: (1) wrongs and (2) rights of redress or recourse. Furthermore, several claims accompany each component.[16] In essence, tort law delineates certain wrongs—assault, defamation, negligence, etc.—and the state empowers victims of those wrongs with a right of action against the wrongdoer. According to Goldberg and Zipursky: "Tort law identifies conduct that is wrongful in the particular sense of being a mistreatment of one by another, and provides recourse through law to the victim against the wrongdoer."[17]

This simple statement requires elaboration because each of the two primary components consists of multiple claims. Goldberg and Zipursky tell us that wrongs are relational guidance rules that are derived from social norms. Furthermore, rights of redress are granted not necessarily to "make whole" but to provide satisfaction to the victim and to preserve civil society.

---

pursued, or perhaps only contemplated, but that need not now exist, or even be attainable. It will often be advanced with improvement of the real world in mind, or even in justification of what is; but it is not necessary that a normative account be at all realistic. In contrast, a descriptive account tells us about existing things in the real world, not how they might be made better, but what they are. Without close attachment to the real world, positive theory has no point.").

15. *See, e.g.*, Goldberg & Zipursky, *Accidents of the Great Society*, *supra* note 3, at 403 ("Our point here is not that [the principle of civil recourse] is demanded by principles of justice, or even morally sound, but that it is the animating idea behind our system of tort law."); *id.* at 364 n.2 ("Most of our work in torts to date has been interpretive."); Zipursky, *Rights, Wrongs*, *supra* note 3, at 6 ("The account I offer is intended to be a framework for a theory of tort law that is descriptive, not prescriptive.").

16. *Cf.* Jane Stapleton, *Evaluating Goldberg and Zipursky's Civil Recourse Theory*, 75 FORDHAM L. REV. 1529, 1529–32 (2006). Stapleton divides civil recourse theory into seven main claims: (1) "A plaintiff must establish all elements of a cause of action before she can ask for a remedy"; (2) "Once these are established, tort law has separate doctrines that determine which remedy will be afforded her"; (3) "[W]e *must* see tort 'obligations' as prospective mandatory directives that enjoin and guide conduct"; (4) We *must* see tort law as relational in analytic structure; (5) "[F]ailure to conform to a tort standard is also judged 'relationally'"; (6) "[W]e must 'reject a reductive-instrumentalist account of duty in terms of the pros and cons of liability rules, and [take] seriously the idea that duty refers to a kind of obligation"; (7) "[I]t provides an account of what is distinctive about the law of torts." *Id.* (emphasis added) (internal quotation marks omitted).

17. Goldberg & Zipursky, *Moral Luck*, *supra* note 3, at 1138; *see, e.g.*, *id.* at 1150 ("[W]e believe that tort liability is predicated on the commission of a wrong—a failure to act in accordance with a relational norm of right conduct—that in turn generates in a victim of the wrong a power to respond to the wrongdoer.").

1. Wrongs

a. Relational

Civil recourse theory is largely a response to what Goldberg and Zipursky refer to as the Holmes-Prosser, or instrumentalist, model of torts.[18] Pursuant to this model, liability is imposed not because a defendant breached an obligation to the injured plaintiff but because the defendant violated public directives, which are influenced by policy considerations and command each and every citizen to refrain from unreasonable conduct that threatens injury to others.[19] In particular, Goldberg and Zipursky take umbrage at Holmes's declaration that torts consist of a duty "'imposed by the State on all the world, in favor of all.'"[20] For Goldberg and Zipursky, nothing less than the essence of tort law is at stake. If Holmes is correct, tort is public law, used to regulate conduct for the public good, not private law, used to resolve disputes between private citizens.

By contrast, civil recourse theory posits tort duties as relational, "owed by specific defendants or classes of defendants to specific plaintiffs or classes of plaintiffs, rather than by each individual to the world at large."[21] Zipursky, in particular, has emphasized relationality, which he refers to as the "substantive standing rule."[22] He argues that relationality explains the New York Court of Appeals's holding in *Palsgraf v. Long Island Railroad Co.*[23] and exists in every tort.[24] Stated differently, relationality means that "a plaintiff does not have a right of action against a defendant on the basis of harm suffered as a result of the defendant's invasion of a third party's right."[25] This is the case because "entitlement to recourse does not spring from the need precipitated by injury. It springs from the affront of being wronged by another."[26]

---

18. Goldberg & Zipursky, *Moral of MacPherson*, *supra* note 3, at 1752–77.

19. *See id.* at 1756–58.

20. *Id.* at 1756 (quoting Oliver Wendell Holmes, Jr., *The Theory of Torts*, 7 AM. L. REV. 652, 661 (1873)).

21. *Id.* at 1744.

22. *See* Zipursky, *Rights, Wrongs*, *supra* note 3, at 15–17.

23. 162 N.E. 99, 100 (N.Y. 1928) (holding that plaintiff has no cause of action because, although defendant breached a duty, it did not breach a duty *to her*).

24. *See* Zipursky, *Rights, Wrongs*, *supra* note 3, at 15–32 (discussing defamation, privacy, fraud, malicious prosecution, false arrest, assault, battery, nuisance, trespass, and emotional and economic harm under the substantive standing rule).

25. *Id.* at 86.

26. *Id.* at 87. For two critiques of the relationality claim, *see* Dilan A. Esper & Gregory C. Keating, *Putting "Duty" in Its Place: A Reply to Professors Goldberg and Zipursky*, 41 LOY. L.A. L. REV. 1225, 1249–71 (2008) and Stapleton, *supra* note 16, at 1544–56.

b. Guidance Rules

These relational wrongs also act as guidance rules, articulating for citizens "how they may and may not treat one another and how they may expect to be treated by others."[27] On this view, tort judgments have an educative mission: "In announcing the kind of conduct that is tortious, courts are enjoining individuals from treating one another certain ways. . . . They serve as norms that guide the behavior of citizens toward one another."[28]

Tort as guidance stands in contrast to the "thin" view of tort judgments as simply "pricing and prohibiting conduct"[29] and Holmes's view that law is aimed at the "bad man."[30] According to Goldberg and Zipursky:

> Law is as much education, explication, articulation, and reinforcement as it is command or threat. . . . Tort law, on this view, is not limited to functioning as a carrot or stick, although it can so function. It does not address the citizen exclusively in his or her capacity as rational maximizer or Holmesian "bad man," although it can do that. In addition, it speaks the language of obligation, helping to settle, as much as possible, what is expected of a person in a range of situations.[31]

In short, tort law as guidance "reinforces and refines norms of responsible conduct."[32]

c. The Source of the Wrongs

If tort is a law of wrongs, from whence come the wrongs? In other words, what is the source of the *content* of the wrongs? According to Goldberg and Zipursky, the legal wrongs comprising tort law are generally

---

27. Zipursky, *Rights, Wrongs*, *supra* note 3, at 5.

28. *Id.* at 92.

29. Goldberg & Zipursky, *Accidents of the Great Society*, *supra* note 3, at 386.

30. O.W. Holmes, *The Path of the Law*, 10 HARV. L. REV. 457, 459 (1897) ("The first thing for a business-like understanding of the matter is to understand its limits, and therefore I think it desirable at once to point out and dispel a confusion between morality and law, which sometimes rises to the height of conscious theory, and more often and indeed constantly is making trouble in detail without reaching the point of consciousness. You can see very plainly that a bad man has as much reason as a good one for wishing to avoid an encounter with the public force, and therefore you can see the practical importance of the distinction between morality and law. A man who cares nothing for an ethical rule which is believed and practised by his neighbors is likely nevertheless to care a good deal to avoid being made to pay money, and will want to keep out of jail if he can.").

31. Goldberg & Zipursky, *Accidents of the Great Society*, *supra* note 3, at 392. For a critique of the guidance rules claim, *see* Stapleton, *supra* note 16, at 1540–44.

32. John C.P. Goldberg, *What Are We Reforming? Tort Theory's Place in Debates over Malpractice Reform*, 59 VAND. L. REV. 1075, 1077 (2006).

context-specific social norms that judges—and occasionally legislatures—elevate to the status of law.[33] These legal wrongs, in turn, have a reciprocal effect on the social norms by the very guidance function discussed above.[34] Thus, legal wrongs are: (1) comprised of social norms; (2) context specific; (3) elevated to legal norms by judges or legislatures; and (4) have a reciprocal effect on social norms.

In general, the legal norms come from social norms. Zipursky notes that legal wrongs are "forms of wrong and are derived from norms that enjoin them and attach to them a species of opprobrium."[35] Specific to tort law, Goldberg refers to these norms as "social norms of safe conduct," "norms of safe practices," and "safety norms."[36] How do Goldberg and Zipursky conceive social norms? To them, they are "obligations already recognized in familiar forms of social interaction."[37] They write that "[i]n different settings and situations, with respect to different sorts of interactions, individuals conceive of themselves as occupying different sorts of normative space governed by different norms of responsibility that impose different sorts of demands or expectations on them."[38] As should be obvious from the description, these norms are sensitive to context; Goldberg has stated that tort law "carves up the social world into 'loci of responsibility'—i.e., particular contexts governed by norms of appropriate conduct that actors must observe for the benefit of identifiable classes of potential victims."[39]

---

33. *See infra* notes 39–42 and accompanying text.

34. Of all the claims of civil recourse theory, this one is covered the least. *See, e.g.*, Alan Calnan, *In Defense of the Liberal Justice Theory of Torts: A Reply to Professors Goldberg and Zipursky*, 1 N.Y.U. J.L. & LIBERTY 1023, 1025 n.16 (2005) ("Where do these norms come from? How do they achieve 'legal' status?"). Most of the following account is taken from Goldberg and Zipursky's direct pronouncements on the subject. However, I extrapolate a few points from Goldberg's discussion of Judge Benjamin N. Cardozo in John C.P. Goldberg, Note, *Community and the Common Law Judge: Reconstructing Cardozo's Theoretical Writings*, 65 N.Y.U. L. REV. 1324 (1990). I do so because Goldberg and Zipursky have explicitly modeled civil recourse theory on what they call "Cardozoan 'pragmatic conceptualism.'" Goldberg & Zipursky, *Unrealized Torts*, *supra* note 3, at 1627, 1628 n.6. Furthermore, analyzing two of Cardozo's major decisions figured prominently in early articles developing the theory. *See* Zipursky, *Rights, Wrongs*, *supra* note 3, at 7–15 (discussing the role of *Palsgraf* in articulating the concept of relationality).

35. Zipursky, *Civil Recourse, Not Corrective Justice*, *supra* note 3, at 746–47.

36. John C.P. Goldberg, *The Constitutional Status of Tort Law: Due Process and the Right to a Law for the Redress of Wrongs*, 115 YALE L.J. 524, 608 (2005).

37. Goldberg & Zipursky, *Accidents of the Great Society*, *supra* note 3, at 392. Goldberg and Zipursky appear to use the term "social norms" in the same way as Professor Cass Sunstein, who defines them as "social attitudes of approval and disapproval, specifying what ought to be done and what ought not to be done." Cass R. Sunstein, *Social Norms and Social Roles*, 96 COLUM. L. REV. 903, 914 (1996).

38. Goldberg & Zipursky, *Accidents of the Great Society*, *supra* note 3, at 392.

39. Goldberg, *supra* note 36, at 608 (citing Goldberg & Zipursky, *Accidents of the*

But how is a social norm elevated to the status of a legal norm? That is the purpose of judges and, less often in tort, legislatures. According to Goldberg, Cardozo's view, which I take to be the idea behind civil recourse theory, is that "the proper function of the law is to articulate and enforce at least some of the obligations recognized in and by the community."[40] Thus, the function of judges within the law is to determine "what portion of the whole spectrum of moral obligations is 'wisely and efficiently enforceable by the aid of jural sanctions.'"[41]

All social norms are not elevated into legal norms, nor should they be. Furthermore, in some cases, legal norms have no correlation in morality:

> Duty in tort law is about legal obligation, and legal obligations are, in many respects, the same sort of creature as moral obligations. Both involve the setting of standards of obligatory conduct; both involve an injunction concerning how to act (particularly with regard to others); both involve social pressure and expectations of a certain kind; both are aimed to preserve important human goods. And while it is not accidental that the two overlap to a considerable extent, it is also not the case that law necessarily derives from or tracks morality. Rather, it is because legal systems and legal obligations are developed with an eye to achieving and safeguarding many of the goods that are also achieved and safeguarded by moral obligations. Nevertheless, because law comes with consequences that morality does not (most obviously state-enforced sanctions), and because there are, at times, demands on law that it take a certain form that renders it efficacious, capable of being internalized, and amenable to application by judges, there will be times at which it is appropriate for legislatures and judges and jurors to decline to elevate certain moral norms to legal norms. Similarly, there are sometimes reasons that favor recognition of legal norms that do not have counterparts in morality.[42]

Finally, the influence of social norms on legal norms is not unidirectional. Legal norms affect social norms as well: "[Tort law] does build on, amplify, and revise obligations that are already recognized, in part because of habits and customs that both shape and are shaped by law."[43] Tort law can be seen "as an effort to recognize, refine, reinforce, and revise

---

Great Society, *supra* note 3, at 403–04).

40. Goldberg, *supra* note 34, at 1335.

41. *Id.* at 1338 (quoting BENJAMIN NATHAN CARDOZO, *The Paradoxes of Legal Science, reprinted in* THE SELECTED WRITINGS OF BENJAMIN NATHAN CARDOZO 273 (M. Hall ed. 1947)). The answer to that jurisprudential question, which is beyond the scope of this article, could fill volumes. Cardozo's immediate answer to the question was to follow the "intelligent and virtuous." CARDOZO, at 274.

42. John C.P. Goldberg & Benjamin C. Zipursky, *Seeing Tort Law from the Internal Point of View: Holmes and Hart on Legal Duties*, 75 FORDHAM L. REV. 1563, 1586–87 (2006).

43. Goldberg, *supra* note 36, at 608.

obligations that are instinct in various standard social interactions."[44] And indeed, at times, the legal obligation will precede and create the social or moral obligation, as Goldberg and Zipursky assert occurred in *Tarasoff v. Regents of the University of California*.[45]

2. Rights of Redress

a. Satisfaction

Once the legal wrong is proved, a right of redress provides the injured victim the ability to act against the injurer and obtain satisfaction. As Goldberg notes, "The animating ideas here are relational and retaliatory, involving notions of empowerment, response, and satisfaction."[46] The idea of satisfaction significantly differs from the traditional understanding of make-whole: "[These ideas] stand in contrast to standard renditions of the make-whole notion, which treat tort law as a means by which a person who suffers a harm can have that harm annulled, erased, or indemnified."[47] The "make-whole" idea that a victim is entitled to "damages equal to the value of past and future economic losses, pain and suffering, and other losses"[48] is a "plausible metric"[49] for damages. However, it is unnecessary. Indeed, tort law offers what Zipursky calls "a diversity of remedies," a point he uses to critique corrective justice theory.[50] What matters is that the remedy provides satisfaction; it must be "such that the victim should reasonably feel that the law has taken her grievance seriously."[51]

---

44. Goldberg & Zipursky, *Accidents of the Great Society*, *supra* note 3, at 391.

45. Goldberg and Zipursky note:

> A range of moral obligations may exist, in some quasi-articulate form, prior to the legal recognition of an obligation, and may be part of the reason for the recognition of such an obligation—or, as in the case of *Tarasoff* and its progeny, tort law may lead the way. But whatever the order, the two levels (and, perhaps, a moral/social/legal spectrum in between) are mutually supportive. Through the law, and in the law, a locus of responsibility is crystallized into a public reality.

Id. at 404 (citing Tarasoff v. Regents of the Univ. of Cal., 551 P.2d 334 (Cal. 1976)).

46. John C.P. Goldberg, *Two Conceptions of Tort Damages: Fair v. Full Compensation*, 55 DEPAUL L. REV. 435, 436 (2006).

47. *Id.*

48. Goldberg & Zipursky, *Moral Luck*, *supra* note 3, at 1142.

49. *Id.*

50. Zipursky, *Civil Recourse, Not Corrective Justice*, *supra* note 3, at 710–13.

51. Goldberg & Zipursky, *Moral Luck*, *supra* note 3, at 1142.

b. Preservation of Civil Society

The rights of redress also help to preserve civil society. Tort "has long empowered individuals and preserved civil society by providing them an avenue for peacefully seeking redress from those who can fairly be held responsible, under law, for having wronged them."[52] The right of redress is, in a sense, a substitute for private revenge: Tort law is available because "those who have been wronged are entitled to some avenue of recourse against the wrongdoer. But, in a civil society, private violence is not permitted, even where there has been a legal wrong."[53] The focus on redress as satisfaction makes sense in light of tort's role as a substitute for private violence. One would forgo such violence if one was "satisfied."

B. The Historical Account

To bolster the account of tort as a law of wrongs, Goldberg, in particular, has turned to history. He does this to establish that tort as wrongs has a long lineage.[54] One obstacle in Goldberg's way is the common assertion by legal historians that tort law developed in the mid-to-late nineteenth century.[55] As an example, Goldberg cites Professor G. Edward White:

> The emergence of Torts as an independent branch of law came strikingly late in American legal history. . . . Torts was not considered a discrete branch of law until the late nineteenth century. The first American treatise on Torts appeared in 1859; Torts was first taught as a separate law school subject in 1870; the first Torts casebook was published in 1874.[56]

Goldberg argues that this view confuses form for substance because of the adoption of the term "torts" in the mid-to-late nineteenth century: "Historians are prone to equate the new use of the term 'torts' with the first attempts by lawyers and scholars to treat personal injury law as a unified, substantive body of law."[57] In reality, Goldberg states, what we now refer to as "torts" in the common law dates back to medieval England.[58] The term "trespass" was used in the thirteenth century as the name of a particular

52. Goldberg, *supra* note 10, at 1519.

53. Goldberg & Zipursky, *Unrealized Torts*, *supra* note 3, at 1643.

54. *See* Goldberg, *supra* note 6, at 1225.

55. *Id.*

56. *Id.* (quoting G. EDWARD WHITE, TORT LAW IN AMERICA: AN INTELLECTUAL HISTORY 3 (exp. ed. 2003)). White's view is widely shared by legal historians. *See, e.g.*, LAWRENCE M. FRIEDMAN, A HISTORY OF AMERICAN LAW 223 (3d ed. 2005); JOHN FABIAN WITT, THE ACCIDENTAL REPUBLIC: CRIPPLED WORKINGMEN, DESTITUTE WIDOWS, AND THE REMAKING OF AMERICAN LAW 7 (2004).

57. Goldberg, *supra* note 6, at 1226.

58. *Id.* at 1227.

chancellor's writ but more generally as "a transgression by one person against another."[59]

Such transgressions could obviously include not only conduct now referred to as torts but also crimes and breaches of contract as well. However, Goldberg chronicles the creation of a distinction between tort and crime and tort and contract, beginning with the former. Pointing to two cases that rely on a distinction between private and public wrongs,[60] Goldberg claims that by 1600 "judges and lawyers understood that a body of law devoted to the redress of personal wrongs—breaches of duties of non-injury owed to others—existed apart from the law of crimes."[61] The distinction became more ingrained as time progressed.[62] Matthew Hale referred to these personal wrongs as "civil wrongs" in his treatise *Analysis of the Law*.[63] Hale contrasts such civil wrongs "'wherein at the Suit or Prosecution of the Party injur'd, he has *Reparation* or *Right* done' with criminal wrongs prosecuted by the Crown."[64]

More significantly, Goldberg points to William Blackstone's reliance upon a distinction between private and public wrongs. Book III of Blackstone's *Commentaries* begins with John Locke's insistence on such a distinction.[65] Misconduct that "'violat[es] . . . public rights and duties, which affect the whole community, considered as a community'" is a public wrong or crime.[66] On the other hand, if the conduct is "'an infringement or privation of the private or civil rights belonging to individuals, considered as individuals,' it is a private wrong or civil injury."[67] Goldberg notes that Blackstone followed Locke in treating the breach of such duty as the deprivation of another's right, leading in tort law to the grant of a remedial privilege to the victim to respond to her injuries.[68] Goldberg further notes that Blackstone's list of personal tort actions—assault, battery, defamation,

---

59. *Id.* (citing S.F.C. MILSOM, STUDIES IN THE HISTORY OF THE COMMON LAW 1–2 (1985)).

60. *See* Goldberg, *supra* note 6, at 1229–31 (discussing EDWARD COKE, 1 THE SELECTED WRITINGS OF SIR EDWARD COKE 481 (Steve Sheppard ed., 2003) (1656)).

61. Goldberg, *supra* note 6, at 1231.

62. *Id.* 1231–32.

63. *Id.* at 1227–28 (citing MATTHEW HALE, THE ANALYSIS OF THE LAW 113 (1713)). Goldberg notes that the book was published posthumously; it was written in the mid-to-late 1600s. *Id.* at 1228.

64. *Id.* at 1228 (quoting HALE, *supra* note 63, at 113).

65. Goldberg, *supra* note 36, at 547 (citing WILLIAM BLACKSTONE, 3 COMMENTARIES 2 (1768)). This is significant because Goldberg explicitly states that civil recourse theory builds on the work of Locke and Blackstone. John C.P. Goldberg, *Wrongs Without Recourse: A Comment on Jason Solomon's Judging Plaintiffs*, 61 VAND. L. REV. EN BANC 9, 13–14 (2008).

66. Goldberg, *supra* note 36, at 547 (quoting BLACKSTONE, *supra* note 65, at 2).

67. *Id.* at 547–48 (quoting BLACKSTONE, *supra* note 65, at 2).

68. *Id.* at 548.

false imprisonment, malicious prosecution, medical malpractice, nuisance, and trespass to land—is recognizable to us today.[69]

Goldberg acknowledges that, for a period, this category of private wrongs encompassed Blackstone's personal tort actions and contract actions. The tort-contract distinction "was less noticeable from within the writ system."[70] Both "were housed under the writ of trespass on the case."[71] The "idea of a relatively complete and freestanding body of contract law only ripened at the turn of the nineteenth century."[72] Goldberg concedes the possibility that, as the field of contracts grew in importance and the writ system collapsed, jurists needed another term to encapsulate the area of private wrongs that was not contracts.[73] "Torts" was the term they selected.[74] Even if this is the case, Goldberg notes that what we call torts today was already "recognized as part of the broad-but-coherent category of private wrongs."[75]

Given this pedigree, Goldberg states that historians who assert that tort had no existence prior to the mid-nineteenth century must be making a different type of claim.[76] It was not that tort did not exist prior to 1850; instead what was known as tort prior to 1850 was "so different in substance or in practice from modern tort law that it is not accurate or helpful to think of modern tort as continuous with its previous incarnations."[77]

At this point, we reach a major point of contention between Goldberg and Zipursky and those holding what they regard as the conventional, academic understanding of tort law. Whereas Goldberg and Zipursky view tort as a law of wrongs, many scholars view tort as accident law—focused on compensation, deterrence, or both—and history plays a large role in that view.[78] As an example, Goldberg offers Professor Lawrence Friedman's thoughts on the increase in personal injuries due to the Industrial Revolution:

> Existing tort law was simply not designed to deal with collisions, derailments, exploding boilers, and similar calamities. . . . Because the job was new, the resulting law was new. There was some continuity in phrasing, but this was in a way misleading. Tort law was new law in the nineteenth century.[79]

---

69. *Id.* (citing BLACKSTONE, *supra* note 65, at 120–28, 138).
70. Goldberg, *supra* note 6, at 1231–32.
71. *Id.* at 1232.
72. *Id.*
73. *Id.*
74. *Id.*
75. *Id.*
76. *Id.*
77. *Id.*
78. *See* discussion *supra* Part II.A.1.
79. Goldberg, *supra* note 6, at 1225–26 (quoting FRIEDMAN, *supra* note 56, at 223).

Under this view, the adoption of the new name "torts" coincided with a new orientation for the underlying law it denoted.[80]

Goldberg argues there are two possible ways to interpret the response to the large increase in accidental injuries during the Industrial Revolution.[81] One version is that the adoption of the term "tort" was part of the "re-conceptualization of the character and purpose" of the non-contractual types of actions formerly known as "private wrongs" or "civil wrongs."[82] Pursuant to this understanding, the flood of accident litigation transformed what formerly had been the adjudication of wrongdoing into a procedure focusing on deterrence, compensation, or both.[83] Policymakers realized that traditional tort law could not effectively handle the wave of accidents, so they altered it.[84] The alternative view is that the wave of accidents "demonstrated decisively the shortcomings of a legal system that responded to accidental injurings exclusively through contracts and torts,"[85] in other words, through agreements about workplace safety and adjudication about whether the employer wrongly caused the accident.[86] Thus, the industrial accidents prompted the design of new laws and institutions—like workers' compensation—to provide greater deterrence and compensation.[87] Under this view, tort as wrongs exists side by side with safety regulations for deterrence purposes, as well as "first-party insurance coverage for expenses related to accidents."[88]

Goldberg argues that the first interpretation is considerably less plausible than the second.[89] Here we finally understand the full importance of history for Goldberg: His proof is largely historical. In determining which of the two interpretations is more convincing, Goldberg asks us to consider whether it was likely that the new "use of the term 'torts' was accompanied by the abandonment of the 500-year-old practice of inviting and adjudicating claims by injury victims against wrongdoers allegedly responsible for those injuries in favor of a new scheme of accident prevention or relief provision."[90] For Goldberg, "[g]iven that the shift in usage left intact all the traditional apparati of tort law—including the concepts used by lawyers and judges to argue about the proper resolution of tort claims—the case for continuity seems vastly stronger than the case for

80. *Id.* at 1236.
81. *Id.* at 1235–36.
82. *Id.* at 1227, 1232, 1235.
83. *Id.* at 1236.
84. *Id.* at 1238–39.
85. *Id.* at 1236.
86. *Id.*
87. *Id.* at 1239.
88. *Id.* at 1240.
89. *Id.* at 1236.
90. *Id.*

discontinuity."[91] Furthermore, although the number of accidental—as opposed to intentional—tort claims skyrocketed during this time period, "accidents often involve[d] wrongful conduct."[92] Because "departures from [the negligence] standard can readily be deemed wrongs in a genuine, non-trivial sense,"[93] tort was, and is, a law of wrongs.

In sum, civil recourse theory is a nuanced attempt to unify tort as a law of wrongs: "[A] plaintiff in a tort case has a right of action *only* because the defendant has committed a legal wrong against the plaintiff, that is, *only* if the plaintiff has suffered a [legal] wrong at the hands of the defendant."[94] Tort law "is a law of wrongs and redress (or recourse)," and "its role in our legal system is not to deliver deterrence and compensation."[95]

III. INSTRUMENTALISM IN TORT LAW

Civil recourse theory must be given its due: it is a comprehensive and subtle attempt to explain the law of torts from an individual justice perspective. At some point, Goldberg and Zipursky have addressed nearly every tort from this perspective.[96] Even one of its most prominent critics has called civil recourse theory a "project of extraordinary breadth and energy."[97] Furthermore, the theory's fundamental proposition—that tort law involves deontic, or duty-based, wrongs—seems unassailable.[98] Civil recourse theory errs, however, when it asserts that tort law is "unitary,"[99] the idea that tort law is *only* about deontic wrongs.[100] The step too far is the

---

91. *Id.*

92. *Id.* at 1237.

93. *Id.* at 1238.

94. Goldberg & Zipursky, *Unrealized Torts*, *supra* note 3, at 1643 (emphasis added).

95. John C.P. Goldberg, Anthony J. Sebok & Benjamin C. Zipursky, *The Place of Reliance in Fraud*, 48 ARIZ. L. REV. 1001, 1014 (2006).

96. *See, e.g.*, Goldberg & Zipursky, *Moral Luck*, *supra* note 3, at 1149–63 (explaining the objectivity of breach and the significance of rights-based forms of strict liability—the over-inclusive side of tort compared to moral obligations); Goldberg & Zipursky, *Moral of MacPherson*, *supra* note 3, at 1799–1811; Goldberg & Zipursky, *Unrealized Torts*, *supra* note 3, at 1672–93 (explaining why tort sets tight limits on liability for nonfeasance and emotional harm—the under-inclusive side of tort compared to moral obligations); Zipursky, *Rights, Wrongs*, *supra* note 3, at 15–32 (discussing defamation, privacy, fraud, malicious prosecution, false arrest, assault, battery, nuisance, trespass, negligence, and emotional and economic harm).

97. Stapleton, *supra* note 16, at 1562.

98. *See id.* at 1556 ("It is obvious, or at least so it seems to me, that, whatever theory-minded academics might like to think, judicial reasoning in tort cases is often quite explicitly couched in non-instrumental terms. There can scarcely be a doubt that we owe moral duties, or that courts take some account of these when they consider whether to recognize a legal obligation . . . .").

99. Goldberg, *supra* note 6, at 1251–52.

100. *Id.* at 1276.

disavowal of compensation, deterrence, and administrative efficiency—which I will refer to collectively as instrumentalism—as part of tort law.

A. The Concessions

During the course of their voluminous scholarship, Goldberg and Zipursky have either admitted or conceded the possibility that some aspects of tort law—including wrongful death and survival actions, some applications of transferred intent, strict liability for abnormally dangerous activities, products liability for manufacturing defects, and punitive damages—are not based on wrongs.

The first concession, for wrongful death and survival actions, appeared in 2003, when Goldberg wrote, "The requirement of having been wronged is not applicable to all tort actions. Legislatures have occasionally relaxed the wronging requirement, as by enacting survival and wrongful death actions that permit certain beneficiaries to sue those who have wronged their decedent."[101] Goldberg then conceded that wronging may not always apply transferred intent: "It is possible, moreover, that the common law also recognizes limited exceptions to the wronging requirement. This may be the case, for example, with respect to certain applications of 'transferred intent.'"[102] For Goldberg, this would "depend on whether the tortfeasor's intentional [actions] toward one person [would] also constitute[]" wrongful actions toward the person actually injured.[103] If the defendant did not have reason to know that the ultimate victim was nearby, the court that invokes the doctrine of transferred intent may be doing so as an act of strict liability or as a waiver of wrongdoing to the plaintiff.[104]

Regarding strict liability for abnormally dangerous activities, Goldberg and Zipursky note that some scholars have "perhaps helped to isolate pockets of truly strict, non-wrongs-based liability that stand in contrast to the general character of tort as a law of wrongs. For example, liability for blasting or other abnormally dangerous activities may not be genuinely wrongs-based."[105] Strict liability for manufacturing defects may also be based in compensation, or loss spreading, as opposed to wrongs.[106]

---

101. John C.P. Goldberg, *Rethinking Injury and Proximate Cause*, 40 SAN DIEGO L. REV. 1315, 1341 n.71 (2003).

102. *Id.*

103. *Id.*

104. *Id.*

105. Goldberg & Zipursky, *Moral Luck*, *supra* note 3, at 1153 n.108; *see also* Goldberg & Zipursky, *supra* note 42, at 1586 n.72 ("It may be that a small subsection of the domain of cases commonly treated as strict liability cases—namely those involving abnormally dangerous activities and wild animals—are instances in which tort law functions as a scheme of liability rules (or as Keeton-esque 'conditional duties').").

106. *See* Goldberg, *supra* note 36, at 613–14. According to Goldberg, it depends on whether Justice Traynor of the California Supreme Court was correct in asserting that loss

Finally, regarding punitive damages, Goldberg and Zipursky acknowledge that "[c]onsiderations of deterrence frequently influence the size of [a punitive damages] award."[107] Zipursky has written an article about punitive damages that he and Goldberg describe as "recognizing respects in which some jurisdictions have invested punitive damages with a significant deterrent role."[108] However, Goldberg and Zipursky do not concede that the point of punitive damages is to deter certain forms of wrongdoing.[109] Instead, they contend that it is "largely" a matter of the "plaintiff's expanded right of individual redress."[110]

Before proceeding, it is important to characterize Goldberg and Zipursky's position in light of their concessions. Is it accurate to claim that they attempt to unify tort law when they concede that aspects of it are or might be based on considerations other than wrongdoing? Goldberg explicitly states, "The requirement of having been wronged is not applicable to all tort actions."[111] Goldberg and Zipursky's extensive body of work as a whole, however, appears to regard tort law as a law of wrongs and the incursions by instrumentalist rationales as *de minimis*. Both before and after the concessions, they repeatedly assert that tort is a law of wrongs.[112] They argue that a plaintiff has a cause of action "*only* because a defendant has committed a legal wrong against the plaintiff."[113] They assert that tort law is "unitary."[114] They assert the role of tort law "is not to deliver deterrence [or] compensation."[115] Although Goldberg and Zipursky have written extensively and exhaustively about tort law for over a decade, they make their concessions in a handful of footnotes that cover a only small percentage of tort claims. Therefore, for Goldberg and Zipursky, the amount of tort claims that are not, or might not, be explicable in terms of wrongdoing is *de minimis*.

However, the admitted existence of instrumentalist rationales in tort law begs two questions: First, how did these rationales come to exist in a body of law allegedly based in wrongdoing? Second, what is the breadth of their influence? Or, to put it in the context of this discussion, is the existence of instrumentalist rationales in tort law really *de minimis*? And

---

spreading was the goal of strict liability for manufacturing defects or whether it was "a means of easing the victim's burden of proving fault." *Id.*

107. Goldberg & Zipursky, *Moral Luck*, *supra* note 3, at 1141 n.58.

108. *Id.* (citing Benjamin C. Zipursky, *A Theory of Punitive Damages*, 84 TEX. L. REV. 105 (2005)).

109. *See id.* ("This is not to say, however, that the point of tort law or even of punitive damages is to deter certain forms of wrongdoing.").

110. *Id.* (citing Zipursky, *supra* note 108).

111. Goldberg, *supra* note 101, at 1341 n.71.

112. *See, e.g.*, *supra* notes 6, 17, 21 and accompanying text.

113. Goldberg & Zipursky, *Unrealized Torts*, *supra* note 3, at 1643 (emphasis added).

114. Goldberg, *supra* note 6, at 1251–52.

115. Goldberg, Sebok & Zipursky, *supra* note 95, at 1014.

finally—to rephrase the question one more time—is tort law, as a positive matter, unified or pluralist?

B. History Revisited

Assume that Goldberg is correct in asserting that what came to be known as tort was a law of wrongs until the mid-nineteenth century.[116] Are there factors supporting the common understanding that the mid-nineteenth through early twentieth century was transformational and that tort was "new law" thereafter? Several events occurred during this time period that, collectively, had the ability to alter the existing law of civil wrongs: (1) the collapse of the writ system; (2) the disentangling of tort and contract; (3) the creation of negligence as an independent cause of action, though not necessarily a tort at the time; (4) the tremendous increase in accidental injuries during the Industrial Revolution; (5) the advent of liability insurance; and (6) the rise of an ethic of interdependence.[117]

1. The Collapse of the Writ System

Until the mid-to-late nineteenth century, "procedural formalities dominated common-law thinking."[118] The various civil-wrong causes of action were funneled through the intricacies of the writ system: "As far as the courts were concerned, rights were only significant, and remedies were only available, to the extent that appropriate procedures existed to give them form."[119] As Professor J.H. Baker explains, "[T]he formulae through which justice was centralised and administered by the king's courts in the twelfth and thirteenth centuries were frozen as part of the 'due process of law' guaranteed by charters of liberties, and gave rise to a formalistic legal culture which affected legal thought at every turn."[120]

In the early parts of the nineteenth century, Jeremy Bentham argued for a "distinction between substantive law and procedure" that would provide an impetus to discard the writ system.[121] This distinction helped Bentham, along with John Austin, to argue that the common law was "intellectually and practically incoherent because *substantive* legal rights and duties were

---

116. *But see* Robert L. Rabin, *The Historical Development of the Fault Principle: A Reinterpretation*, 15 GA. L. REV. 925, 927–28 (1981) (arguing that the pre-industrial era was substantially governed by a "no-liability" approach; thus, it is important to consider both what was included in and what was excluded from its parameters when analyzing tort law).

117. *See infra* Part.III.B.1–6.

118. J.H. BAKER, AN INTRODUCTION TO ENGLISH LEGAL HISTORY 53 (4th ed. 2002).

119. *Id.*

120. *Id.*

121. *See* Thomas C. Grey, *Accidental Torts*, 54 VAND. L. REV. 1225, 1231 n.10 (2001) (citing JEREMY BENTHAM, *Principles of Judicial Procedure with the Outlines of a Procedure Code*, *in* 2 THE WORKS OF JEREMY BENTHAM 1, 5 (John Bowring ed., 1962)).

learned and classified for practice under the jumbled array of *procedural* forms that had grown up over the centuries to enforce them."[122] According to Bentham this was backwards: "[P]rocedure should be designed" to aid substance.[123]

Based on Bentham's principles, the New York legislature proposed the "Field Code" in 1848.[124] The Code abolished the writs in favor of a "unitary 'civil action,' under which plaintiffs were simply to plead facts that established grounds for the relief sought."[125] The Code "served as a kind of catalytic agent for procedural reform elsewhere in the United States."[126] By 1876, twenty states had adopted "some version of the reformed procedure."[127]

The forms of action had provided the common law with a structural underpinning; causes of action were arranged according to the procedural forms.[128] The law needed reconceptualization once this underpinning was removed, as the collapse of the forms of actions triggered "a juristic debate about the taxonomic arrangement of the substantive law."[129] Professor G. Edward White noted that, during the same period, an "independent impetus in the rise of university-based legal education" arose with a concomitant "interest of a new group of scholars in law as an autonomous subject inviting systematic conceptual elaboration."[130]

2. Disentangling Contract

Early in the search for the substantive categories that would comprise private law, "contracts" was seen as a fitting organizational structure: "[F]rom about 1850 on, English and American legal writers came to agree that contracts would be one fundamental branch of the new substantive private law."[131] Though common law contract actions first emerged around 1600, "the idea of a relatively complete and freestanding body of contract law only ripened at the turn of the nineteenth century."[132] Although a

---

122. *Id.* at 1240.

123. *Id.*

124. *Id.* at 1231.

125. *Id.*

126. LAWRENCE M. FRIEDMAN, A HISTORY OF AMERICAN LAW 391 (2d ed. 1985). Friedman's chapter "Procedure and Practice: An Age of Reform" provides a history of the events leading to the Field Code as well as the spread of procedural reform across the United States. *Id.* at 391–411.

127. Grey, *supra* note 121, at 1231 n.8. England passed reform statutes in 1852 (Common Law Procedure Act) and 1873 (Judicature Act). *Id.*

128. *Id.* at 1231.

129. *Id.* at 1226–27.

130. *Id.* at 1232 n.11 (citing WHITE, *supra* note 56, at 3–19 (with particular emphasis on pp. 8–11)).

131. *Id.* at 1235–36.

132. Goldberg, *supra* note 6, at 1232.

consensus developed concluding that the relatively nascent field of contracts should be a building block of civil law, torts did not enjoy the same legitimacy: As late as 1871, Oliver Wendell Holmes, Jr. stated that "[t]orts is not a proper subject for a law book."[133]

Though Holmes later changed his mind, and torts, like contracts, became a building block of civil law, the two categories had to be disentangled. Under the writ system, the main action that would become "contracts" and most of the actions that would become "torts" were filed as a trespass on the case.[134] There was substantial overlap between the two substantive categories, and thus, difficulty in drawing a strict dichotomy between them.[135] In his 1887 treatise on torts, Frederick Pollock stated that "the attempt to classify personal actions as arising in either contract or tort could not 'be defended as a scientific dichotomy.'"[136] Furthermore, Pollock noted that the distinction created "'considerable perplexity'" in the large "'intersection between the two regions.'"[137] Thus, large portions of law were left "arbitrarily classifiable as either contract or tort."[138] Quoting F.W. Maitland, Professor Thomas Grey writes that "'the courts of the present day are very free to consider the classification of causes of action without paying much regard' to the tort-contract dichotomy."[139]

3. The Creation of Negligence as an Independent Cause of Action

It was into this confused atmosphere that the independent tort of negligence was born. Though negligence has what Professor Grey refers to as "a long prehistory,"[140] until the mid-nineteenth century there was no "general purpose" tort of negligence.[141] Grey notes that negligence "emerged from the action of trespass on the case to take fairly clear shape in a few prescient judicial decisions and bits of commentary in the 1860s."[142]

---

133. Oliver Wendell Holmes, 5 AM. L. REV. 340, 341 (1871) (book review); *see also* Grey, *supra* note 121, at 1262 & n.109.

134. Grey, *supra* note 121, at 1251. *See generally* D.J. IBBETSON, A HISTORICAL INTRODUCTION TO THE LAW OF OBLIGATIONS 126–51 (1999) (discussing the development of the action of assumpsit).

135. *See* Grey, *supra* note 121, at 1252 ("[T]he ambiguous character of large bodies of important doctrine counseled against any program of sorting civil claims into categories of contract and tort.").

136. *Id.* (quoting FREDERICK POLLOCK, THE LAW OF TORTS 431–32 (1887)).

137. *Id.* (quoting POLLOCK, *supra* note 136, at 431–32).

138. *Id.* at 1251 (citing POLLOCK, *supra* note 136, at 431–32).

139. *Id.* at 1250 (quoting F.W. Maitland, *Historical Note on the Classification of the Forms of Personal Action*, *in* FREDRICK POLLOCK, THE LAW OF TORTS 370 app. A (1887)).

140. *Id.* at 1260.

141. KENNETH S. ABRAHAM, THE LIABILITY CENTURY: INSURANCE AND TORT LAW FROM THE PROGRESSIVE ERA TO 9/11 20 (2008).

142. Grey, *supra* note 121, at 1260.

However, even as the independent negligence cause of action was being created, its relationship to the law of torts was uncertain. In the 1860s and 1870s, treatise writers were divided not only over whether negligence was an independent cause of action but also whether it was a part of the nascent law of torts.[143] Grey notes that these divisions were understandable, given that negligence was "mainly considered . . . not as a distinct tort cause of action, but as an element of liability in a variety of civil actions ranging across a number of fields of law."[144] Until that time, "common-law commentary" had chiefly treated negligence as a part of bailments, which scholars considered either a "separate legal category in itself or a subdivision of property or contract law."[145] Of course, the overwhelming majority of modern tort claims in the United States are based in negligence.[146] Thus, in the 1860s and 1870s, both whether the heart of modern tort law was an independent cause of action and even whether it was a tort at all was debatable.

In 1873, Oliver Wendell Holmes, Jr. announced his revised view that torts was indeed a proper subject and helped to tie negligence to the law of torts in his article *The Theory of Torts*.[147] In the article, Holmes organized tort law around the concept of negligence, construed as an objective, "public standard of reasonable behavior."[148] In his tripartite classification of torts (which is still in use today), Holmes centered negligence between two extremes: strict liability and intent.[149] Comprised "of both moral fault and no-fault [elements], objective negligence mediated between the two poles of the tort spectrum."[150] Holmes's negligence analysis, which formed the basis for his discussion of torts in *The Common Law*,[151] had enormous influence over the development of tort law.[152]

4. The Surge in Accidental Injuries During the Industrial Revolution

As the negligence cause of action was emerging, the scale of its potential application was expanding dramatically. After the Civil War, the increased rate of industrialization, the growth of railroads, and the spread of

143. *Id.* at 1260–62.

144. *Id.* at 1262.

145. *Id.*

146. *See* STEVEN K. SMITH ET AL., BUREAU OF JUSTICE STATISTICS, U.S. DEP'T OF JUSTICE, NCJ-153177, TORT CASES IN LARGE COUNTIES: CIVIL JUSTICE SURVEY OF STATE COURTS, 1992 (1995).

147. Holmes, Jr., *supra* note 20, at 659–62.

148. Grey, *supra* note 121, at 1266–67.

149. *Id.* at 1270.

150. *Id.*

151. HOLMES, *supra* note 2, at 63.

152. *See* Grey, *supra* note 121, at 1269. It is no wonder that Goldberg and Zipursky, in attempting to remedy what they see as the academy's current view of torts, single Holmes out for criticism; *supra* note 23 and accompanying text.

the population across the continent augmented the risk of injury as people went about their lives, particularly in the workplace.[153] Professor John Fabian Witt has described the results: "Industrializing economies in the mid to late nineteenth century experienced an explosion of accident rates alongside the rapid development of new industries and more powerful machinery."[154] The accident rates due to industrialization came to dwarf even those of war.[155] In 1891, the New Jersey Bureau of Statistics of Labor and Industries stated that "'the destruction of human life . . . in the peaceful pursuits of industry . . . [was] far greater than during the four years of destruction in the late civil war.'"[156]

On a national level, the 1850 census was the first to tabulate deaths from accident.[157] Between 1850 and 1880, "the [percentage] of deaths attributable to accident[s] among men aged ten to fifty increased by over 70 percent, [rising] from 7 percent to 12 percent."[158] Of course, some industries were more dangerous than others. From 1860 to 1890, the percentage of deaths from railroad accidents among males aged ten to fifty increased nearly fivefold, from 0.6 to 3%.[159] Similarly, the increase in accident rates was worse in certain regions, such as the northeast.[160] One study of Philadelphia found "that the accidental death rate . . . [rose] from 34.4 accidental deaths per 100,000 [people] between 1839 and 1845, to 58.6 accidental deaths per 100,000 people between 1895 and 1901."[161] Compared to modern accident rates, those around the turn of the twentieth century were stunning: In 1912, a study of accidental deaths "estimated 82,500 deaths per year; [since that time], the population of the United States has tripled, but the number of accidental deaths has increased by less than a quarter."[162]

---

153. ABRAHAM, *supra* note 141, at 26–27.

154. WITT, *supra* note 56, at 22.

155. *Id.* at 24 (quoting BUREAU OF STATISTICS OF LABOR & INDUS. OF N.J., THIRTEENTH ANNUAL REPORT 367 (1891)).

156. *Id.* (quoting BUREAU OF STATISTICS OF LABOR & INDUS. OF N.J., THIRTEENTH ANNUAL REPORT 367 (1891)).

157. *Id.* at 26 (citing J.D. DE BOW, SUPERINTENDENT, U.S. CENSUS, MORTALITY STATISTICS OF THE SEVENTH CENSUS OF THE UNITED STATES, 1850 17–20 (Washington, D.C., A.O.P. Nicholson 1855)).

158. *Id.*

159. *Id.* (citing JOHN S. BILLINGS, DEP'T OF THE INTERIOR, REPORT ON VITAL AND SOCIAL STATISTICS IN THE UNITED STATES AT THE ELEVENTH CENSUS: 1890, PART I.—ANALYSIS AND RATE TABLES 740–45 (Washington, D.C., Gov't Printing Office 1896)); *see also* SEC'Y OF THE INTERIOR, STATISTICS OF THE UNITED STATES 1860 53–55 (Washington D.C., Gov't Printing Office 1886).

160. WITT, *supra* note 56, at 26.

161. *Id.* (citing ROGER LANE, VIOLENT DEATH IN THE CITY: SUICIDE, ACCIDENT, AND MURDER IN NINETEENTH-CENTURY PHILADELPHIA 36 (Ohio State Univ. Press 2d ed. 1999) (1979)).

162. *Id.* at 26–27.

By far the worst category of accidental injuries was workplace injuries, which "represent[ed] close to one-third of all accidental deaths and . . . between one-half and two-thirds of all accidental injuries."[163] "In 1890, railroad worker death rates were 314 per 100,000 workers per year."[164] The same year, coal mining deaths "rang[ed] from 215 deaths per 100,000 workers per year in bituminous coal mines to 300 deaths per 100,000 workers per year in anthracite coal mines."[165] Trainmen, who "operat[ed] the coupling devices between [rail]cars, and brakemen, who operated the train's handbrakes, died in work-related accidents at rates of 900 and 1,141 deaths per 100,000 workers per year, respectively."[166] "Industrialization," Witt summarized, "had devised . . . new and unfamiliar mechanisms for inflicting harm on the human body."[167]

A rise in litigation paralleled the rise in accidents. Between 1870 and 1890, "the number of accident cases being litigated in New York City's state courts grew almost eightfold; by 1910 the number had grown again" more than five times.[168] From 1870 until 1910, the percentage of tort cases in New York City trial courts' contested caseload rose from 4.2 to 40.9%.[169] Boston courts experienced similar pressure from increased volume.[170] As late as 1880, no more than "a dozen or so suits were filed in [Boston's] superior court alleging damage caused by negligent operation of a horsecar."[171] By 1900, "there were over 800 [accident] cases involving streetcars in superior court, and 600 more in the municipal court."[172]

5. The Advent of Liability Insurance

In part to deal with the increase in the accident rate and concomitant tort liability, demand rose for liability insurance.[173] However, the common law was ambivalent about liability insurance's validity into the 1880s. As Professor Kenneth Abraham has explained, liability insurance "was a lawful transaction that produced social benefits"; however, it also created a "moral hazard."[174] In other words, insurance could encourage people to be

---

163. *Id.* at 27.

164. *Id.*

165. *Id.*

166. *Id.*

167. *Id.* at 28.

168. *Id.* at 59.

169. WITT, *supra* note 56, at 59 (citing RANDOLPH E. BERGSTROM, COURTING DANGER: INJURY AND LAW IN NEW YORK CITY, 1870–1910 20 tbl. 4 (1992)).

170. *See* ROBERT A. SILVERMAN, LAW AND URBAN GROWTH: CIVIL LITIGATION IN THE BOSTON TRIAL COURTS, 1880–1900 105 (1981).

171. *Id.*

172. WITT, *supra* note 56, at 59 (citing SILVERMAN, *supra* note 170, at 105).

173. ABRAHAM, *supra* note 141, at 20.

174. *Id.*

less cautious by covering the costs of their negligence.[175] The most significant issue—whether insurance protecting people from their own negligence was valid—was heavily influenced by the United States Supreme Court's 1886 ruling in *Phoenix Insurance Co. v. Erie Transportation Co.*[176] In *Phoenix*, the Court noted that an insured, by purchasing insurance, "does not diminish his own responsibility to the owners of the goods, but rather increases his means of meeting that responsibility."[177] Simply put, the increase in moral hazard was acceptable if insurance also compensated victims of negligent loss.[178]

Although *Phoenix* was not itself a case about liability insurance, the Court's opinion seemed to signal that such insurance was acceptable. In the spring of 1886, the first liability insurance company in the United States, Employers' Liability Assurance, opened an office in Boston and began selling policies.[179] As the name of the company suggests, the demand for liability insurance was initially concentrated on employers' liability for workplace accidents.[180] However, insurers were soon branching out. Coverage for automobile liability appeared soon after the first automobiles were produced in the late nineteenth century.[181] Although originally sold as part of a policy covering horse-drawn coaches and carriages, by 1905 automobile liability insurance was sold in separate policies.[182] By 1909, twenty-seven liability insurance companies were operating in the United States, "enough to warrant publication of a manual to assist in establishing rate classes and fixing premiums."[183]

6. An Ethic of Interdependence

The last factor, more intangible than those preceding it, is the Progressive Era moderation of the rugged individualism that persisted in the United States in the nineteenth century. Professor G. Edward White highlighted what he calls an ethic of "social interdependence" in the early twentieth century as part of his intellectual history of tort law, stating, "In the early years of the twentieth century a vision of society as a constellation of interdependent groups displaced a competing vision . . . of society as an

175. *Id.*

176. 117 U.S. 312, 324 (1886).

177. *Id.*

178. ABRAHAM, *supra* note 141, at 26.

179. *Id.* at 28. Ten years later, the Maryland Court of Appeals rejected the first direct challenge to liability insurance. *Id.* (citing Boston & A.R. Co. v. Mercantile Trust & Deposit Co. of Baltimore, 34 A. 778, 787 (Md. 1896)).

180. *Id.*

181. *Id.* at 70–71.

182. *Id.* at 71.

183. *Id.* at 32–33.

aggregate of autonomous individuals."[184] A distinction of the ethic of interdependence was a sense of collective responsibility for social problems. For instance, the "problems of disadvantaged groups" were seen as "affecting all of society; they were problems of social living, not of individual character."[185] In the Progressive Era, "[p]overty, unemployment, adverse working conditions, child labor, and industrial injuries came to be perceived as phenomena for which society bore some collective responsibility rather than simply as costs of the struggle of life."[186]

Based on the foregoing factors, how might tort be considered "new law" after the middle of the nineteenth century? First, the cause of action that dominates modern tort law, negligence, was created thereafter.[187] Second, the focus of modern tort law, accidents, exploded just as the negligence cause of action was created.[188] Thus, tort operated with a new cause of action for a type of conduct that had not often been the subject of "tort" actions.[189] Until the middle of the nineteenth century, what we know as tort law primarily dealt with intentional acts, such as assault and battery.[190]

Assuming tort was a law of wrongs prior to the middle of the nineteenth century, the significant question for purposes of this article is how these factors might have altered its character. At a minimum, the shift in focus from intentional to negligent conduct,[191] based on an objective standard, is a move from greater to lesser blameworthiness. Does it go further? The negligence cause of action was created at a time of categorical confusion.[192] Tort and contract were being sorted through, with many of the doctrines "arbitrarily classifiable" as either.[193] In fact, the nature of negligence itself was contested; at one time, it was primarily seen as a bailments doctrine, grading the degree of care required by the bailee.[194] This intermixture with contract may have adulterated the once-pure law of tort as wrong.

In addition, the nation's ethos shifted in the early twentieth century. In the Progressive Era, the idea of interdependence became more widely accepted, encroaching into the rugged individualism that had dominated the nineteenth century.[195] Pursuant to an interdependent view, problems are

---

184. WHITE, *supra* note 56, at 66.

185. *Id.*

186. *Id.* at 69–70.

187. *See supra* Part III.B.3.

188. *See supra* Part III.B.4.

189. *See supra* Part III.B.3–4.

190. *See* G. Edward White, *The Intellectual Origins of Torts in America*, 86 YALE L.J. 671, 690 (1977).

191. *See id.* at 689.

192. *See id.*

193. *See supra* note 138 and accompanying text.

194. *See supra* note 145 and accompanying text.

195. *See* WHITE, *supra* note 56, at 66.

seen as a collective, as opposed to an individual, responsibility.[196] In other words, social norms shifted away from an individualistic and toward a collective focus.[197] It was through this collective focus that the explosion in accidents during the Industrial Revolution was seen.[198] The advent of liability insurance provided a pragmatic means to allow a collective attitude to alter tort law.[199]

Goldberg and Zirpursky have acknowledged or addressed most of the six aforementioned factors.[200] Goldberg offers the most direct response to the challenge that tort was new law and not a law of wrongs in his 2008 Monsanto lecture, *Ten Half-Truths About Tort Law*.[201] Recall that Goldberg distinguishes two versions of the claim.[202] The first is that there was "a re-conceptualization of the character and purpose of civil litigation brought by injury victims against alleged injurers . . . such that what looked by all appearances to be adjudications of claims of wrongdoing actually was not."[203] However, "[a] second and distinct claim is that the era of industrial accidents demonstrated decisively the shortcomings of a legal system that responded to accidental injurings exclusively through contracts and torts . . . ."[204] In other words, any loss-allocation or deterrence elements of law would be applied outside of tort. Goldberg presents the choice starkly: Either "tort was quietly, nearly secretly, transforming itself into a law of loss-allocation or deterrence" or "tort law continued to function as a law of wrongs and redress."[205]

However, there is a third possibility between the extremes that Goldberg offers. Tort law did not transform root and branch into a system of instrumentalism concerned with loss-allocation and deterrence. Nor did it remain a pure system of deontological wrongs. Instead, while retaining its wrongs-based character, tort was infused with elements of instrumentalism.

---

196. *Id.*

197. *Id.*

198. *See supra* Part III.B.4.

199. This sketch does not contradict Professor Calnan's historical account of early torts as wrongs based. *See* ALAN CALNAN, A REVISIONIST HISTORY OF TORT LAW 205 (2005). Calnan's history of tort as wrongs does not cover the late nineteenth and twentieth centuries. *Id.* Furthermore, Calnan acknowledges the move from "trespass by force and arms" to trespass on the case was a move from greater to lesser wrongdoing. *Id.* at 204–05. Most significantly, Calnan expresses skepticism about these later developments, arguing that they tend to corrupt and degrade the law's core principles and values. *Id.* at 293–94.

200. *See, e.g.*, Goldberg, *supra* note 6, at 1231–32 (collapse of the writ system), 1232 (disentangling contract), 1234–40 (Industrial Revolution explosion of accident cases), 1264–70 (insurance); Goldberg & Zipursky, *Moral of MacPherson*, *supra* note 3, at 1746 n.45 (creation of negligence as an independent tort).

201. Goldberg, *supra* note 6, at 1240.

202. *See supra* notes 81–88 and accompanying text.

203. Goldberg, *supra* note 6, at 1235–36.

204. *Id.* at 1236.

205. *Id.* at 1238.

Of course, this hypothesis—and the role of the six factors—is simply conjecture absent evidence within the doctrines and application of tort law itself. It is to this evidence that I now turn, bearing in mind that Goldberg and Zipursky have themselves pointed to instances of loss-allocation or deterrence in tort law.

C. Doctrine Revisited: Tort Reform

In this subsection, I examine tort reform during the last century to demonstrate the growth of instrumentalism in tort law. The focus on tort reform in no way exhausts the instrumentalism in tort law,[206] but it suffices to demonstrate the pluralism of tort law from a positive perspective. Because many tort reforms are statutory, it is important to note at the outset that Goldberg and Zipursky accept that statutes are a legitimate part of tort law.[207] I will examine tort reforms chronologically:[208] (1) workers'

206. In addition to the concessions that Goldberg and Zipursky have made, several tort causes of action, or elements thereof, have been potentially influenced by instrumentalism. They include trespass (an essentially strict liability tort originally designed to avoid breaches of the peace), nuisance (often balancing gravity of harm and utility of conduct), conversion (another essentially strict liability tort), and vicarious liability. Furthermore, each element of a negligence cause of action arguably includes instrumentalism. Duty is often presented as a multi-factor policy analysis, which includes factors like concern over "crushing exposure to liability." Strauss v. Belle Realty Co., 482 N.E.2d 34, 36 (N.Y. 1985). Even if the "Hand Test" is not used to determine breach, social "utility" is often considered part of the analysis. *See* RESTATEMENT (SECOND) OF TORTS § 291 cmt. d (1964). Recent innovations to cause-in-fact analysis—loss of chance, alternative liability, and market share liability—are at least partially based in policy considerations. Proximate cause (or "scope of liability") is often shaped by courts using the vision of "public policy" and "practical politics" from Judge Andrews's dissent in *Palsgraf.* Palsgraf v. Long Island R. Co., 162 N.E. 99, 103 (N.Y. 1928) (Andrews, J., dissenting). Finally, damages are frequently capped for instrumentalist efficiency reasons. *See infra* Part III.C.4. Furthermore, Jane Stapleton notes that "there are areas of tort law that can only be accounted for in instrumental terms, for example torts that are explicitly based on the violation of some public policy such as the tort of retaliation by an employer against an employee." Stapleton, *supra* note 16, at 1531. Indeed, Goldberg served as a judicial clerk to Judge Jack Weinstein, whom he described as having a Jamesian (compensation-oriented) approach to mass tort litigation. John C.P. Goldberg, *Misconduct, Misfortune, and Just Compensation: Weinstein on Torts*, 97 COLUM. L. REV. 2034, 2037 (1997). As is obvious, to do justice to the subject, a separate article would be necessary. I thank Alan Calnan and Jane Stapleton for their contributions to this list.

207. *See* Goldberg, *supra* note 6, at 1271 ("At the most basic level, the point to be made is the obvious one that statutes figure in tort law in all sorts of ways. Indeed, it is difficult to think of an aspect of tort law that has not been touched by statutory law.").

208. There is some overlap between reform of automobile accidents and products liability. Efforts to reform automobile accidents began considerably earlier, and I treat them first.

compensation; (2) automobile accidents; (3) products liability; and (4) "modern" tort reforms.[209]

The progression from near total separation of instrumentalism and tort law in workers' compensation to tort law's embrace of instrumentalism in modern tort reforms is striking. Workers' compensation was adopted as a *substitute* for tort law.[210] Those advocating its adoption contrasted the instrumentalist compensation scheme with tort law's strict emphasis on fault.[211] The two were seen as very different; in fact, one of the chief selling points of workers' compensation was its contrast with tort law.[212] Twenty years later, when the focus was on automobile accidents, the separation between the instrumental aim of reform and tort law was not as stark.[213] There was a sense that the fault standard in tort had been relaxed somewhat.[214] Furthermore, the no-fault automobile reforms ultimately enacted were integrated into the tort system in a way that workers' compensation simply was not.[215] The next major reform, products liability, was based in not only the instrumental aim of compensation, but also the instrumental aim of deterrence.[216] Unlike the preceding tort reforms, this was a formal inclusion of instrumentalism within tort law itself.[217] However, it was restricted to a relatively small set of cases.[218] The final set of tort reforms, such as damage caps and collateral source reform, have been enacted across many jurisdictions.[219] Largely based in the instrumental aim of administrative efficiency, the reforms apply to large swaths of tort cases.[220] All of the reforms are based in instrumentalism.[221] As time progressed, the reforms shifted from external to tort law, to partially external to tort law, to internal to tort law but limited in scope, and finally to internal to tort law and widely applicable.[222]

---

209. My purpose here is not to advocate for or against any of the reforms. Some I support; others I oppose. Instead, my goal is to demonstrate the increasing instrumentalism in tort law through the progression of major tort reforms over the course of the last century.

210. *See* Christopher Howard, *Workers' Compensation, Federalism, and the Heavy Hand of History*, 16 STUD. IN AM. POL. DEV. 28, 31–32 (2002).

211. *See id.* at 32.

212. *See id.*

213. *See infra* Part III.C.2.

214. *See infra* Part III.C.2.

215. *See infra* Part III.C.2.

216. *See infra* Part III.C.3.

217. *See infra* Part III.C.3.

218. *See infra* Part III.C.3.

219. *See infra* Part III.C.4.

220. *See infra* Part III.C.4.

221. *See infra* Part III.C.4.

222. *See infra* Part III.C.4.

1. Workers' Compensation

The first major (and still the most thorough) "tort reform" was workers' compensation.[223] For present purposes, two points about workers' compensation are significant: First, it was premised on the instrumentalist rationale of compensation.[224] Second, it was regarded as separate from tort law.[225]

Concern over supporting families in the absence of a provider led to criticism of the tort system's treatment of workplace accidents.[226] As noted earlier, industrialization created a surge in workplace injuries, many of which were serious.[227] Recall that between 1850 and 1880, the percentage of deaths attributable to accidents in males aged ten to fifty increased by over 70%.[228] The vast majority of accidental injuries occurred in the workplace.[229] During that time period, the vast majority of paid workers were men.[230] Based on the ideology of the day, many men were sole wage-earners for their families.[231] Many victims injured in workplace accidents lost the ability to provide for their families for some period of time.[232] Increasingly they turned to the tort system for relief.[233] As previously mentioned, between 1870 and 1890, "the number of accident cases being litigated in New York City's state courts grew almost eightfold; [and] by 1910 the number had grown again" more than five times.[234] The problem was that workers seldom recovered under tort law; instead, they bore the full costs of accidents and were often forced into poverty.[235]

This "compensation gap" sparked criticism of tort law from multiple sources. In a January 1908 message to Congress, President Theodore Roosevelt stated that the great "outrage" of work accidents under the common law of torts was that "the entire burden of the accident falls on the helpless man, his wife, and his young children."[236] Many in the organized bar shared Roosevelt's opinion; a 1913 publication of the Ohio State Bar

223. *See* Howard, *supra* note 210, at 28.
224. *See id.* at 32.
225. *See id.*
226. *See id.* at 31.
227. *See supra* Part III.B.4.
228. WITT, *supra* note 56, at 26.
229. *Id.* at 27.
230. *See id.* at 131–32.
231. *See id.* at 131.
232. *See id.* at 130.
233. *Id.* at 51–52.
234. *See supra* note 168 and accompanying text.
235. *See* Howard, *supra* note 210, at 31.
236. 2 Theodore Roosevelt, *The Employers' Liability Law, in* THE ROOSEVELT POLICY 699, 700 (Current Literature Publ'g Co. 1908)).

Association stated that "a system of laws which permit[s] no recovery in so large a percentage of deaths and injuries occurring is unjust."[237]

The most influential criticism of contemporary tort law came from attorney Crystal Eastman in her book *Work-Accidents and the Law*.[238] The book, written as part of the Pittsburgh Survey's study of social conditions in the "nation's most important industrial city,"[239] was "'perhaps the strongest single force in attracting public opinion'" to the work accident problem.[240] Eastman examined 526 separate cases of industrial deaths in Allegheny County, Pennsylvania during portions of 1906 and 1907.[241] Virtually all fatal workplace accidents, 523 of 526, killed males.[242] Almost half were married men, and 63% were the sole supporter of their family.[243] Eastman concluded that "[t]he people who perished were those upon whom the world leans."[244] In only 30% of the cases did the family receive more than $500, which was approximately one year of income for the lowest paid of the deceased workers.[245] Eastman then recounted the fate of the widows and orphans, complete with poignant photographs of their plight.[246] For example, one widow and her children were forced from their home into the back rooms of a parent's house.[247] Eastman concluded that the common law approach to workplace accidents, which often left the injured man and his family to bear the most of the loss in most cases, was unjust.[248] Ultimately, her criticisms of tort law led her to recommend workers' compensation.[249]

In the wake of Eastman's book, state and federal commissions charged with studying workplace accidents "also made dependent wives and children central objects of concern."[250] Indeed, Professor John Fabian Witt credits Eastman with "organiz[ing] work-accident debates around the image of the wounded family."[251] In the months following the publication of her

---

237. WITT, *supra* note 56, at 43.

238. CRYSTAL EASTMAN, WORK-ACCIDENTS AND THE LAW (Paul Underwood Kellogg ed., 1910).

239. WITT, *supra* note 56, at 126.

240. *Id.* at 129 (quoting Roy Lubove, *Workmen's Compensation and the Prerogatives of Voluntarism*, 8 LAB. HIST. 254, 255 (1967)).

241. EASTMAN, *supra* note 238, at 119.

242. *Id.* ("Of the 526 people killed, 258, almost one-half, were married men; 265 were single men or boys . . . .").

243. *Id.* at 119–20.

244. *Id.* at 119.

245. *Id.* at 122.

246. *Id.* at 137.

247. *Id.*

248. *Id.* at 131.

249. *Id.* at 220.

250. WITT, *supra* note 56, at 131.

251. *Id.* at 130.

book in 1910, support for workers' compensation spread like a "prairie fire" or a "whirlwind."[252]

Eastman's fame from her research on workplace accidents led to an appointment on the Wainwright Commission in New York State (known formally as the New York State Employers' Liability Commission), which drafted the first workers' compensation statute enacted in the United States.[253] The commission also focused on supporting dependents after a workplace accident.[254] The first of four reasons it listed for dissatisfaction with the common law system was "[t]hat only a small proportion of the workmen injured by accidents of employment get substantial compensation, and, therefore, as a rule, they and their dependents are forced to a lower standard of living and often become burdens upon the State through public or private charity."[255] Its study of New York State accidents found that only 54% of the families surveyed received some form of insurance benefit from the death of a male wage-earner.[256] Moreover, 60% of those receiving the benefit received less than $500.[257] Families of injured or killed workmen "must depend upon the work of women and children or upon the assistance of relatives and friends, must reduce their standard of living to the detriment of health, and must often become destitute and dependent upon charity."[258] In recommending workers' compensation, the Commission stated that the amount of compensation to be provided to a workman should "keep him and those dependent on him out of absolute destitution."[259]

In its analysis, the Commission went beyond the need of families and laid the groundwork for the compensation theory of tort law. Most significantly, the Commission discussed the idea of loss or risk spreading: "Though the workman cannot shift this accident burden upon the cost of the product or upon the trade, the employers can through their power to fix the selling price of the product."[260] It is a point the Commission repeated: "[P]art of the burden of these inevitable accidents which now rests upon the workmen, least fitted of all to carry it, should be shared by those who profit by such work, that is, by the employer, who ultimately will shift that burden

---

252. *Id.* at 127 (internal quotation marks omitted).

253. *Id.*

254. *Id.*

255. N.Y. STATE EMP'RS LIAB. COMM'N, REPORT TO THE LEGISLATURE OF THE STATE OF NEW YORK BY THE COMMISSION APPOINTED UNDER CHAPTER 518 OF THE LAWS OF 1909 TO INQUIRE INTO THE QUESTION OF EMPLOYERS' LIABILITY AND OTHER MATTERS 19 (1911) [hereinafter WCR].

256. WITT, *supra* note 56, at 99 (citing WCR, *supra* note 255, at 26).

257. *Id.*

258. WCR, *supra* note 255, at 27.

259. *Id.* at 50.

260. WCR, *supra* note 255, at 7; *see* ABRAHAM, *supra* note 141, at 40 ("It was in debates about workers' compensation that the first sustained, though comparatively primitive, arguments about risk spreading and enterprise liability were made.").

to the consumer."[261] The second point involves the nature of accidents. If accidents tend to be caused by bad behavior, a fault-based system is defensible. If, on the other hand, accidents are inevitable, a more compensation-oriented theory is preferable. The Commission reported that statistics demonstrate that "there is a mathematical ratio of industrial accidents in the hazardous trades, depending on the speed at which industry moves and the number of workmen, as remorseless and as certain as the death rate on which the tables for life insurance are based."[262]

Of course, Goldberg is aware of the surge of accidents following the Industrial Revolution and the criticisms of tort law it engendered, that tort was poor accident law or that it prompted a transformation of tort law from a law of wrongs and redress into a law of compensation and deterrence.[263] Instead, he argues that "the accident epidemic prompted the realization that the branch of the law devoted to the enterprise of redressing wrongs will not always or even usually be as efficacious as certain alternatives in reducing the number of accidents or in reliably getting prompt relief to accident victims."[264] According to Goldberg, the government is within its rights to supplement or even replace tort law in certain areas, such as

---

261. WCR, *supra* note 255, at 48; *see also id.* at 39 ("[T]hey have accepted the principle that each industry should be made to bear the burden of its personal accident losses, in the same way that it already bears the burden of accidental losses to plant and machinery. The employer is selected to act as the agent in adding the cost of workers' compensation for industrial accidents to the other costs of production, because that is the simplest and most direct way of accomplishing the desired result. It is assumed that he will be reimbursed for this expense, as for his other expenses of production, in the prices he receives for his products."), 67 ("If the bridge-builder in the State of New York must pay more to his employees in the shape of damages or compensation, that additional cost will be reflected in the total cost of building the bridges, just as would a rise in wages or a rise in the cost of materials; but when the law is known the bridge-builder and the house-builder will make his contracts accordingly and his prices accordingly, and no man will be deprived of property or unduly muleted. The community at large will then support the injured workman by compensation through the employer in the first instance, rather than through increased taxes or charity.").

262. WCR, *supra* note 255, at 5. The committee's brief reference to deterrence does not fit easily with its emphasis on the inevitability of accidents. *Id.* at 7 ("We have been impressed by the fact that employers generally (there are many exceptions) pay less attention to prevention of accidents than the public interest demands because the payment for the damages of accident bears very little direct relation under the present system of liability, to the number of accidents and we hope and believe that the changes in the liability laws we recommend, because they tend to make the employer pay something for every accident, will have a real effect in making him put his mind constantly on the question of preventing accidents."). On the importance of statistics as an approach to accidental injuries, see WITT, *supra* note 56, at 138–51.

263. Goldberg, *supra* note 6, at 1236.

264. *Id.* at 1239.

workplace accidents, to better achieve compensation and deterrence.[265] That does not mean, however, that *tort* aims to compensate or deter.[266]

At least regarding the enactment of workers' compensation, Goldberg is correct about the separation of tort law from the instrumentalist rationale of compensation.[267] First, and most significantly, workers' compensation was adopted as a substitute for tort law.[268] The workers' compensation system removed injuries from the tort system and provided compensation in its place, largely on the basis of causation alone.[269] Second, at the time workers' compensation was enacted, those studying tort law believed it to be founded upon fault.[270] For example, the Commission stated: "The New York system of liability is, speaking generally, founded on fault. . . . That is the fundamental principle of our law, inherited from the common law of England, which no statute in this State has ever changed."[271] By contrast, greatly reducing the role of fault in workplace injuries was one of the crucial attractions of workers' compensation.[272] Therefore, the first major tort reform of the last century, workers' compensation, does not offer direct evidence of instrumentalism operating within tort law.

---

265. Goldberg, *supra* note 32, at 1079. However, Goldberg believes the government's ability to alter tort law is limited to some extent by the Constitution. *See* Goldberg, *supra* note 36, at 626–27.

266. Goldberg, *supra* note 32, at 1079.

267. Although perhaps not rising to the level of integration between instrumentalism and tort, one facet of workers' compensation is worth noting. Workers' compensation did not completely preempt tort law. In many jurisdictions, intentional torts, the most egregious cases, remained viable tort causes of action. In other words, workers' compensation recognized a difference among previously existing tort cases. Some were fit for treatment as a compensable event; others were eligible for continued treatment as "wrongs." Doctrinally, this distinction among tort claims was generally accomplished in two ways. The most common holding was that intentional torts were not "accidents" as many jurisdictions required. *See, e.g.*, Boek v. Wong Hing, 231 N.W. 233, 234 (Minn. 1930). Alternatively, the language in certain compensation statutes explicitly excluded intentional torts from coverage or allowed the employee the choice of compensation or tort law. *See, e.g.*, Workmen's Compensation Act, 1915 Colo. Sess. Laws 515. Eventually forty-one of the fifty-one jurisdictions recognized some form of the intentional tort exclusivity exception to workers' compensation. *See* 6 LARSON'S WORKERS' COMPENSATION LAW §103.01 n.4 (1999) ("Jurisdictions that do not recognize this exception to the exclusiveness found in their workers' compensation statutes include Alabama, Georgia, Indiana (Occupational Disease Act), Maine, Nebraska, New Hampshire, Pennsylvania, Rhode Island, Virginia, and Wyoming.").

268. *See* WCR, *supra* note 255, at 8 (describing in positive terms "the *substitution* for the present vicious system, a new system of workers' accident compensation") (emphasis added).

269. *See id.*

270. *See id.* at 10.

271. *Id.*

272. *See* Howard, *supra* note 210, at 32.

However, the debate over workers' compensation heralded the beginning of tort law's loss-spreading rationale. It was built on the need to support families in the absence of a provider and expanded into arguments over the inevitability of accidents and the ability to spread losses.[273] Even though this rationale was used to remove claims from the tort system in the case of workers' compensation, it would later be used within tort itself.[274]

2. Automobile Accidents

Reform efforts based on compensatory principles were anticipated almost immediately after the states first enacted workers' compensation statutes.[275] In many ways, the history of reform efforts for automobile accidents mirrors the advent of workers' compensation. Along with workplace accidents, there was a sharp increase in automobile accidents combined with congestion in the courts.[276] Just as with workplace accidents, reform proposals were based on the instrumentalist rationale of compensation.[277] However, the reform of automobile accidents was managed in a subtly different way than reform of workplace accidents. The distinction between tort law and the instrumentalist reforms was not as stark.

The late 1910s through the early 1930s witnessed a sharp increase in automobile accidents. From 1915–1932, automobile accidents "multiplied over seven times."[278] Furthermore, in 1932 an author noted the "automobile fatality rate ha[d] increased more than 500% since 1913, while the death rate for other kinds of accidents show[ed] a decline of over 30% for the same period."[279]

With the increase in accidents came reliance on the tort system to cover the losses and, thus, congestion in the courts. Thirty percent of all new cases on the Supreme Court of New York County's calendar between October 1928 and April 1930 were automobile accident cases.[280] Furthermore, a study of the Court of Common Pleas of Philadelphia County indicated that half of all cases tried to a jury were automobile accidents.[281]

---

273. *See id.* ("Employers would pay not because they were negligent, but because accidents were inherent in an industrial society and therefore a cost of doing business.").

274. *See supra* Part III.C.2.

275. *See* Jeremiah Smith, *Sequel to Workers' Compensation Acts*, 27 HARV. L. REV. 235, 235 (1914).

276. *See* ROBERT E. KEETON & JEFFREY O'CONNELL, BASIC PROTECTION FOR THE TRAFFIC VICTIM: A BLUE PRINT FOR REFORMING AUTOMOBILE INSURANCE 11–13 (1965).

277. *See id.* at 7.

278. REPORT BY THE COMMITTEE TO STUDY COMPENSATION FOR AUTOMOBILE ACCIDENTS TO THE COLUMBIA UNIVERSITY COUNCIL FOR RESEARCH IN THE SOCIAL SCIENCES 17 (1932) [hereinafter COLUMBIA PLAN].

279. *Id.*

280. *Id.* at 20.

281. *Id.*

Reform efforts reached a crescendo in 1932 with the "Columbia Plan," a proposal created by the Committee to Study Compensation for Automobile Accidents, a group of academics, lawyers, and social scientists under the auspices of Columbia University.[282] Similar to Crystal Eastman's focus on the wounded family in *Work-Accidents and the Law*, the Columbia Plan covered the plight of accident victims' families,[283] as illustrated by its statement that "this report will be concerned with ways in which families met the accident situation."[284] The committee also emphasized that wage-earners were disproportionately victims of automobile accidents.[285] Yet again, those injured were "those upon whom the world leans."[286] Also reminiscent of Eastman's approach to workplace accidents, the committee included numerous individual "case studies" that humanized the effect of automobile accidents.[287]

However, the Columbia Plan committee was more explicit in its concerns about compensation. Not only did the authors acknowledge their compensatory focus, they disclaimed the instrumentalist approach of deterrence: "The problem of compensation for injuries caused by such accidents rather than the problem of accident prevention has been the Committee's field of study."[288] In the summary, the compensation problem was described as follows:

> Injury or death in a motor vehicle accident means economic loss to the person injured and to the dependents of one who is killed. The incidence of this loss presents the problem of motor vehicle accident compensation. If the persons injured or the families of those who are killed receive no compensation for their economic loss, or if they receive compensation which does not cover their loss, or if the compensation is too long delayed, all or part of the burden is borne by them, and in many cases by their doctors, hospitals, landlords and tradesmen. If adequate and prompt compensation is received, the full burden is borne by the motorist or by his insurer.[289]

The Committee stated in another section: "The problem of compensation is concerned with the distribution of the burden of this loss."[290] Finally, and most significantly, the Columbia Plan committee provided an outline that noted: "[T]he main purpose of [the] compensation

---

282. *See id.* at 2–3.

283. *See, e.g.*, COLUMBIA PLAN, *supra* note 278, at 222.

284. *Id.*

285. *See id.* at app. 260 (Earners were 52% of victims. The next highest category, children, were only 33%. Housewives followed with 13%).

286. EASTMAN, *supra* note 238, at 119.

287. *See* COLUMBIA PLAN, *supra* note 278, at 58–59.

288. *Id.* at 1 (footnote omitted).

289. *Id.* at 199–200.

290. *Id.* at 19.

plan is to spread through insurance the inevitable losses due to automobile accidents."[291]

The Columbia Plan also questioned the principle of fault: "The Committee believes that the principle of liability for fault only is a principle of social expediency, and that it is not founded on any immutable basis of right."[292] Furthermore, even though the law provides a cause of action for damages arising from an automobile accident "only when someone has been at fault,"[293] the Committee acknowledged that juries have a lot of discretion: "In this sense, negligence is what the jury says is negligence."[294] Even more significant than the practice of juries was the practice of insurance adjusters: "[I]n practice, damages are usually paid by insurance carriers without strict reference to the principle of fault."[295]

Based on data from case studies, the Committee concluded, "[I]nsurance companies pay in so large a proportion of the cases in which liability insurance is carried, that the principle of liability without fault seems almost to be recognized."[296] For example, the Committee investigated 2,500 closed cases of temporary disability involving insured defendants and 900 cases without insured defendants.[297] Claimants had received money in 86% of the insured cases but in only 27% of the uninsured cases.[298] For cases of permanent disability, the Committee studied 192 closed cases with insured defendants and ninety without an insured defendant.[299] Claimants received money in 96% of the insured cases but in only 21% of the cases without an insured defendant.[300]

However, the Columbia Plan never achieved sufficient political support for any jurisdiction to enact it.[301] World War II consumed national attention, and automobile accident reforms were placed on the backburner.[302]

In the 1960s, however, hope for reform of automobile accident recovery reemerged. In 1965, Robert E. Keeton and Jeffrey O'Connell published *Basic Protection for the Traffic Victim: A Blueprint for Reforming Automobile Insurance*,[303] which included a draft statute ready to

---

291. Young B. Smith, *Compensation for Automobile Accidents: The Problem and Its Solution*, 32 COLUM. L. REV. 785, 799 (1932).

292. COLUMBIA PLAN, *supra* note 278, at 212.

293. *Id.* at 25.

294. *Id.* at 26.

295. *Id.* at 200.

296. *Id.* at 203.

297. *Id.*

298. *Id.* at 204.

299. *Id.*

300. *Id.*

301. *See* Joseph A. Page, *Roscoe Pound, Melvin Belli, and the Personal-Injury Bar: The Tale of an Odd Coupling*, 26 T.M. COOLEY L. REV. 637, 668 (2009).

302. *Id.*

303. KEETON & O'CONNELL, *supra* note 276.

be introduced by state legislators.[304] Keeton and O'Connell emphasized the continued increase in automobile accidents: "In 1963 the death toll reached a new high of 43,600, which was in turn eclipsed by a figure of about 47,000 to 48,000 in 1964."[305] Furthermore, they predicted the problem would only intensify as the number of cars and the number of younger, more dangerous drivers increased.[306] Moreover, court congestion caused by these accidents continued.[307] Keeton and O'Connell described the effect of victims' attempts to gain compensation for traffic accidents on the courts as "crushing."[308]

Keeton and O'Connell proposed that states adopt "no-fault" automobile laws in which first-party insurance for economic loss caused by personal injuries would be mandatory.[309] In other words, up until a certain monetary threshold—say $10,000—a driver's own insurer would pay the driver for economic loss for personal injuries suffered in automobile accidents, regardless of fault.[310] Eventually, sixteen states adopted some form of no-fault automobile law in the 1970s.[311]

Keeton and O'Connell based no-fault insurance on compensatory, or loss-spreading, principles.[312] They concluded that neither fault nor deterrence were wholly adequate reasons to shift losses.[313] Keeton and O'Connell explained that liability insurance decreases the significance of fault in tort law through settlement practices.[314] Insurers impersonally appraise claims "by standards appropriate to the management of a large pool of risks."[315] Insurers settle an individual claim "whenever this can be done for a sum representing an appropriate discount from the probable amount of an award if the case should be tried and lost. This discount is tailored to the degree of likelihood that the insurer would win if the claim were litigated."[316] The goal is to settle advantageously more than

304. *See id.* at 6–10.

305. *Id.* at 11 (footnotes omitted).

306. *See id.* at 12.

307. *Id.* at 13.

308. *Id.*

309. *Id.* at 7.

310. *See id.*

311. *See, e.g.*, Paul J. Barringer et al., *Administrative Compensation of Medical Injuries: A Hardy Perennial Blooms Again*, 33 J. HEALTH POL. POL'Y & L. 725, 732 (2008). Of the sixteen states, twelve retain some version of no-fault laws today. *Id.*

312. KEETON & O'CONNELL, *supra* note 276, at 249.

313. *Id.*

314. They noted statistics, similar to those relied upon by the Columbia Plan committee, which showed the percentage of victims recovering when insurance is available. *Id.* at 254 n.14 (citing 2 HARPER & JAMES, TORTS 781 (1956)) (87% of persons suffering serious personal injury received some money if insurance was involved).

315. *Id.* at 254.

316. *Id.*

disadvantageously over the course of the entire set of claims.[317] According to Keeton and O'Connell, this creates a system in which fault, or lack thereof, takes a secondary position to resolving claims in an efficient and profitable manner.[318]

Thus, Keeton and O'Connell argued that "the burden of a minimum level of protection against measurable economic loss" should "be treated as a cost of motoring."[319] Specifically, they contended that "[t]he cost of providing this minimum level of compensation for traffic victims would be distributed generally among the persons who benefit from motoring, without regard to fault in particular accidents."[320] This distribution based on a class of persons—drivers—avoids the offensive search for a deep pocket involved in distributing losses among individuals.[321] Keeton and O'Connell noted that this type of loss distribution over a class of persons was already operating within the tort system in the form of liability insurance.[322] Tort liability insurance "distributes losses of a prescribed type among the members of a large class of persons whose conduct creates risks of such losses. Thus, to recognize the legitimacy of tort liability insurance is implicitly to approve this principle of distributing losses among a class."[323]

Thus, the impetus for both workers' compensation and automobile accident reform was a surge in injuries that led to a desire for greater compensation for victims. Compared with workers' compensation, however, the automobile accident context contained less of a sense of separation between fault-based tort and compensatory principles, both as a matter of practice and of doctrine:

> In its practical operation in each state today, the automobile claims system is not one in which compensation is conditioned upon nonfault of the victim and fault of the tortfeasor. Rather, the theory of full compensation or none yields to the practice of partial compensation in almost every one of the multitude of settlements.[324]

Of this "part-recovery-most-of-the-time matter,"[325] Professor Witt remarked, "[T]he ostensibly individualized common law of torts had become a system for the aggregate resolution of personal injury claims—a system that socialized the risks of the activities out of which it arose."[326]

---

317. *Id.*
318. *Id.*
319. *Id.* at 268.
320. *Id.*
321. *Id.* at 249.
322. *Id.* at 250.
323. *Id.*
324. *Id.* at 254.
325. JOHN FABIAN WITT, PATRIOTS AND COSMOPOLITANS 271 (2007).
326. *Id.* at 272. Aiding instrumentalist compensatory principles in the resolution of automobile accidents as a matter of practice were several doctrines created just for that

In addition to less separation between the practice of tort law and the instrumentalist aim of compensation compared to workers' compensation, the reform ultimately enacted exhibited less separation as a matter of doctrine.[327] Workers' compensation was a full-scale substitution of tort law for accidental workplace injuries.[328] In comparison, the Keeton and O'Connell reform proposal leading to the enactment of no-fault automobile laws envisioned a role for compensatory principles *in conjunction with* tort law.

---

purpose. Two prominent examples of these trends are compulsory automobile insurance and expansion of vicarious liability.

Compulsory automobile insurance was a legislative attempt to further the compensation of automobile accident victims. *See* KEETON & O'CONNELL, *supra* note 276, at 251. The statistical chasm between the percentage of victims compensated by insured drivers and those compensated by uninsured drivers led to pressure in some jurisdictions to mandate automobile insurance. *See* Barringer, *supra* note 311, at 732; Gary T. Schwartz, *Auto No-Fault and First-Party Insurance: Advantages and Problems*, 73 S. Cal. L. Rev. 611, 623 (2000). Mandatory insurance plays a different role than voluntary insurance. Voluntary liability insurance is perfectly consistent with an individual justice theory of tort law. *See* Goldberg, *supra* note 206, at 2039 (footnote omitted). One may purchase liability insurance to meet one's responsibilities to others in the event of a judgment. The purchase of insurance means that a potential tortfeasor can balance the moral scales under a corrective justice approach or provide satisfaction under a civil recourse approach. *See id.* Compulsory automobile insurance is different: "The purpose of the compulsory motor vehicle insurance law is not, like ordinary insurance, to protect the owner or operator alone from loss, but rather is to provide compensation to persons injured through the operation of the automobile insured by the owner." Wheeler v. O'Connell, 9 N.E.2d 544, 546 (Mass. 1937). Such laws could have a significant impact. At the time of the Columbia Plan, only one jurisdiction—the Commonwealth of Massachusetts—had adopted compulsory automobile insurance. KEETON & O'CONNELL, *supra* note 276, at 76. Massachusetts achieved nearly 100% insurance rate of resident motor vehicles. COLUMBIA PLAN, *supra* note 278, at 45. By contrast, even including Massachusetts, the committee estimated that 27.3% of private passenger and commercial vehicles registered in the United States carried liability insurance. *Id.*

Another attempt, both by legislatures and the judiciary, to find a financially solvent defendant was the expansion of vicarious liability. By common law, this expansion was largely achieved through the "family automobile doctrine." COLUMBIA PLAN, *supra* note 278, at 28. In essence, courts in about half the states held owners liable for the automobile accidents of members of their family who were using the automobile with consent. *Id.* The Columbia Plan committee described expansion of vicarious liability by legislation as being "in the same direction and . . . significant." *Id.* The committee noted that in at least eight states "an owner is made liable by statute for injury caused by the negligent operation of his motor vehicle by a person who drives with his knowledge and consent, even though the operator is neither his servant nor a member of his family." *Id.* (footnotes omitted). This, of course, had the effect of providing a greater number of insured defendants in automobile accident cases.

327. This cannot be said for the Columbia Plan, which mirrored worker's compensation faithfully.

328. *See* Howard, *supra* note 210, at 31–32.

The Keeton and O'Connell proposal integrated tort and compensatory principles for cases of accidental injury. The basic concept was to cover relatively modest losses on a compensatory basis while retaining tort in cases of severe injury:

> If tort damages for pain and suffering would not exceed $5,000 and other tort damages, principally for economic loss, would not exceed $10,000, an action for basic protection benefits replaces any tort action against a basic protection insured; in cases of more severe injury, the tort action is preserved, but the recovery is reduced by these same amounts.[329]

Keeton and O'Connell saw a synergy in compensatory treatment of smaller cases (and the initial losses of larger cases), along with full-scale tort claims for larger cases. They argued that a distinctive wastefulness arose in applying the fault principle to small cases.[330] A multitude of small claims deprived the seriously injured of prompt trials and fair payments by clogging the courts and draining insurance payments.[331] Furthermore, in larger cases the prompt payment of the first $10,000 for out-of-pocket losses as they accrued would strengthen a plaintiff's bargaining position.[332] No longer as desperate for immediate cash, plaintiffs could reject low-ball settlement offers.[333]

However, there was still a role for fault. Indeed, Keeton and O'Connell conceded that the concept of fault is "deeply rooted in our society and will not be lightly cast aside."[334] However, their proposal retained fault for more than pragmatic, political considerations. Especially in the severe cases, Keeton and O'Connell leave to the tort system accidents where fault is clear, stating that "there is much to be said for visiting all that tort damages entail on the person at fault, and thereby including in the award compensation for pain and suffering accompanying a prolonged and bitter convalescence or permanent disability."[335] Severe automobile accidents would better justify the time and expense required to determine fault.[336] Furthermore, one of the problems with applying the fault standard was its attenuation due to the desire to provide basic compensation to automobile accident victims.[337] If juries knew that basic economic losses were covered in most cases (due to the no-fault provisions), they would better apply the fault standard.[338]

---

329. KEETON & O'CONNELL, *supra* note 276, at 7.
330. *Id.* at 270.
331. *Id.*
332. *Id.* at 269–70.
333. *Id.*
334. *Id.* at 271.
335. *Id.*
336. *Id.* at 272.
337. *Id.* at 271.
338. *Id.* at 272.

The Keeton and O'Connell reform proposal has been described as hybrid,[339] modified,[340] and mixed[341]—all of them emphasizing the mixture of no-fault (compensatory) and tort (fault) principles.[342] The no-fault laws actually adopted by the states divided cases into less severe claims appropriate for compensation and more severe claims appropriate for tort resolution.[343] In doing so, a "threshold" was set, beyond which the claim was considered sufficiently severe to merit tort treatment.[344] The states adopted two different types of thresholds: monetary and verbal.[345] Under a monetary threshold, like the $10,000 in the original Keeton and O'Connell proposal,[346] the damages had to reach a certain dollar value before a tort claim was possible.[347] Under a verbal threshold, the victim must suffer a certain condition, which the statute describes, to pursue a tort claim.[348]

The reform of automobile accidents was based on the same instrumentalist compensatory principles as workers' compensation. However, the strict separation of fault-based tort law from the compensation-based workers' compensation was relaxed somewhat in the automobile context.[349] Both the practice of automobile accident litigation at the time of the reforms and the content of the actual reforms demonstrated a stronger connection between tort and compensation.[350] Regarding practice, the presence of automobile insurance adjusters tended to diminish the importance of fault in resolving automobile injury claims.[351] Regarding doctrine, the reforms adopted in the states combined no-fault compensatory principles for smaller cases with fault-based tort law for more severe cases.[352] Yet there was no acknowledgment that instrumentalist principles were formally operating in tort law. That step occurred in products liability.

---

339. *See, e.g.*, Schwartz, *supra* note 326, at 617.

340. *Id.* at 617 n.22.

341. *Id.*

342. *Id.* at 617.

343. *Id.*

344. *Id.*

345. *Id.*

346. KEETON & O'CONNELL, *supra* note 276, at 7.

347. *See* Schwartz, *supra* note 326, at 617. The monetary thresholds actually adopted tended to be lower than $10,000. For example, Massachusetts, the first jurisdiction to enact a no-fault automobile law, adopted a $2,000 threshold. *See id.* at 618.

348. *Id.* at 617. Michigan, New York, and Florida are verbal threshold states. *Id.* at 617 n.23.

349. *See supra* notes 329–48 and accompanying text.

350. *See supra* notes 329–48 and accompanying text.

351. *See supra* notes 329–48 and accompanying text.

352. *See supra* notes 329–48 and accompanying text.

3. Products Liability

From at least the time of the Columbia Plan through the mid-1950s, tort scholarship focused on the problem of automobile accidents.[353] By contrast, injuries caused by products received little attention.[354] That changed around 1957.[355]

Products liability doctrine began to change soon after it received academic attention,[356] and the changes were instrumentalist. For years, products cases had been handled pursuant to both contract-based warranty and tort-based negligence causes of action.[357] In 1960, the Supreme Court of New Jersey announced its landmark decision in *Henningsen v. Bloomfield Motors, Inc.*[358] *Henningsen* applied warranty without privity to perhaps the most significant product of all, the automobile.[359] Furthermore, even though the holding was based on the implied warranty of merchantability,[360] other contract limitations were invalidated.[361] The instrumentalist reasoning in many places is stark. For example, privity is eliminated for the twin instrumentalist rationales of loss spreading and deterrence: "In that way the burden of losses consequent upon use of defective articles is borne by those who are in a position to either control

---

353. George L. Priest, *The Invention of Enterprise Liability: A Critical History of Modern Tort Law*, 14 J. LEGAL STUD. 461, 463 (1985).

354. *Id.*

355. *Id.* at 504.

356. In 1955, Fleming James, Jr., a tireless advocate of loss spreading and tort liability as "social insurance," set his sights on products liability. *See* Fleming James, Jr., *Products Liability*, 34 TEX. L. REV. 44 (1956). James's treatise, written with Fowler V. Harper, was published one year later. F. HARPER & F. JAMES, JR., THE LAW OF TORTS (1956). Many of the reviews focused on Harper and James's treatment of products liability. *See* Priest, *supra* note 353, at 504 (citing W. Page Keeton, 45 CALIF. L. REV. 230, 232–33 (1957) (book review); Robert A. Leflar, 32 N.Y.U. L. REV. 1156, 1157 (1957) (book review); Warren A. Seavey, 66 YALE L.J. 955, 956–58 (1957) (book review)). In 1960, William Prosser published a highly influential article on strict liability. William L. Prosser, *The Assault upon the Citadel (Strict Liability to the Consumer)*, 69 YALE L.J. 1099 (1960). In the article, Prosser noted a "trend" of jurisdictions extending the strict liability of warranty without privity of contract to products beyond the traditional category of food. *Id.* at 1110–14. In addition, he argued that warranty had a history as a tort as well as a contract doctrine. *Id.* at 1126. Thus, the standard of strict liability for products could and should be used as a matter of tort instead of contract. *Id.* at 1127. This would eliminate the contractual obstacles to recovery, such as privity and the use of warranty disclaimers.

357. *See* James, *supra* note 356, at 44.

358. 161 A.2d 69 (N.J. 1960).

359. *Id.* at 83 ("We see no rational doctrinal basis for differentiating between a fly in a bottle of beverage and a defective automobile.").

360. *Id.* at 84.

361. For instance, privity of contract was invalidated. *Id.* Furthermore, a disclaimer of the warranty was also held to be invalid. *Id.* at 95.

the danger or make an equitable distribution of the losses when they occur."[362]

Henningsen set the stage for the ultimate use of strict liability in tort that Prosser advocated.[363] Professor Priest noted that, despite its use of the implied warranty of merchantability, the case "repudiates every other principle of contract law potentially applicable to product defect actions."[364] Thus, *Henningsen* demonstrated that there was no truly consensual, contractual basis for implied warranty, marking "the effective end of the relevance of contract law in defective product actions involving personal injury."[365]

In 1963, the California Supreme Court decided *Greenman v. Yuba Power Products, Inc.*,[366] which completely seized for tort law the basis of claims for injuries from defective products.[367] After echoing the *Henningsen* court's rejection of contract-based warranty doctrines limiting recovery in personal injury actions,[368] the court shifted the basis of recovery to tort, stating that "[a] manufacturer is strictly liable in tort when an article he places on the market, knowing that it is to be used without inspection for defects, proves to have a defect that causes injury to a human being."[369] Justice Roger Traynor, who authored the opinion, had been trying to establish strict liability in tort for personal injuries caused by products since his concurrence in *Escola v. Coca Cola Bottling Co.*[370] Traynor did not provide an extensive rationale for the decision. Indeed, the only direct statement of rationale in *Greenman* is one of loss spreading: "The purpose of such liability is to insure that the costs of injuries resulting from defective products are borne by the manufacturers that put such products on the market rather than by the injured persons who are powerless to protect themselves."[371]

Justice Traynor, however, had already addressed the reasons for imposing strict liability on the manufacturer at length in his *Escola* concurrence, which he cited in *Greenman*.[372] In the *Escola* concurrence,

---

362. *Id.* at 81.

363. *See* Prosser, *supra* note 356.

364. Priest, *supra* note 353, at 507.

365. *Id.*

366. 377 P.2d 897 (Cal. 1963).

367. *Id.* at 900.

368. *Id.* at 899–900.

369. *Id.* at 900.

370. 150 P.2d 436, 440 (Cal. 1944).

371. *Greenman*, 377 P.2d at 901. Kenneth Abraham states, "The search for categories of defendants that could spread losses broadly, either by insuring or by passing liability costs through to a large customer base, was highly influential in the development of modern products liability." ABRAHAM, *supra* note 141, at 139.

372. As Priest noted, "The power of the *Greenman* opinion is through its reference to Traynor's concurring opinion in *Escola*." Priest, *supra* note 355, at 441. *See Greenman*, 377 P.2d at 901 (citing Justice Traynor's concurrence in *Escola*).

Traynor argued for "absolute liability" based on a combination of the familiar instrumentalist rationales of loss spreading and deterrence.[373] In fact, Traynor started with deterrence: "[P]ublic policy demands that responsibility be fixed wherever it will most effectively reduce the hazards to life and health inherent in defective products that reach the market. It is evident that the manufacturer can anticipate some hazards and guard against the recurrence of others, as the public cannot."[374] However, because not all injuries could be prevented, strict (or absolute) liability would also help spread the losses associated with those injuries: "The cost of an injury and the loss of time or health may be an overwhelming misfortune to the person injured, and a needless one, for the risk of injury can be insured by the manufacturer and distributed among the public as a cost of doing business."[375]

In 1964, Prosser, serving as the Reporter for the *Restatement (Second) of Torts* and emboldened by *Greenman*, orchestrated the passage of Section 402A.[376] This section extended strict liability to sellers of defective and unreasonably dangerous products regardless of fault. [377] Over the next several years, jurisdictions steadily adopted strict liability for products.[378] To such jurisdictions, the "powerful arguments of [*Henningsen* and *Greenman*] provided the grounds."[379] These arguments, and thus the rationales supporting the adoption of strict liability, were instrumentalist.

---

373. *See* Escola, 150 P.2d at 440–44.

374. *Id.* at 440–41.

375. *Id.* at 441. This combination of compensation and deterrence rationales was also known as "enterprise liability." For the history of enterprise liability, *see, e.g.*, Priest, *supra* note 353; John Fabian Witt, *Speedy Fred Taylor and the Ironies of Enterprise Liability*, 103 COLUM. L. REV. 1, 3 (2003) (tracing the central ideas of enterprise liability to the "efforts of late nineteenth- and early twentieth-century engineers to remake the firm").

376. *See* Priest, *supra* note 353, at 512–14.

377. RESTATEMENT (SECOND) OF TORTS § 402A (1965). For a discussion of its passage, *see* Priest, *supra* note 353, at 518.

378. Priest, *supra* note 353, at 518.

379. *Id.* at 507. A quick glance at several jurisdictions supports Priest's thesis that jurisdictions adopted strict liability for instrumentalist reasons. *See, e.g.*, Santor v. A & M Karagheusian, Inc., 207 A.2d 305, 311–12 (N.J. 1965), *abrogated by* Alloway v. Gen. Marine Indus., 695 A.2d 264 (N.J. 1997) (citing *Henningsen*, *Greenman*, and Prosser, *supra* note 356, at 1124–34) (referring to the manufacturer's obligation as "an enterprise liability" and stating that "[t]he purpose of such liability is to insure that the cost of injuries or damage, either to the goods sold or to other property, resulting from defective products, is borne by the makers of the products who put them in the channels of trade, rather than by the injured or damaged persons who ordinarily are powerless to protect themselves."); First Nat'l Bank in Albuquerque v. Nor-Am Agric. Prods., 537 P.2d 682, 695 (N.M. Ct. App. 1975) (citing Prosser, *supra* note 356, at 1122–23) ("Allowing injured plaintiffs to proceed on a theory of a manufacturer's liability, without the necessity of proving negligence, will cause manufacturers to take cautionary steps to prevent the marketing of dangerously defective products. Such preventive measures may avert tragedies . . . and thereby save our system the cost of lawsuits . . . ."); Leichtamer v. Am. Motors Corp., 424 N.E.2d 568, 575

Although instrumentalist rationales were explicitly invoked in the reform of products liability, such cases are a limited subset of torts. Furthermore, the later development of products liability doctrine reduced the "strictness" by which liability was determined. When drafting Section 402A, it appears the paradigmatic defect in the minds of its drafters was a manufacturing defect, one in which a product departed from its intended design and caused injury.[380] Litigation in the 1970s uncovered two additional types of defects: design and warning.[381] Unlike manufacturing defects, design and warning defects affect every unit of a product.[382] In design defects, the shortcoming is in the way the entire line of products was planned.[383] In warning defects, the manufacturer provided incorrect or insufficient information about the safe use of a product.[384] Liability for manufacturing defects is indeed strict.[385] On the other hand, liability for design defects is "ordinarily . . . a question of weighing the costs and benefits of that design and any reasonable alternatives to it."[386] Liability is found for warning defects "only if there was 'inadequate' disclosure of the risks posed and those risks could have been reduced by 'reasonable' warnings or instructions."[387] As Abraham stated, these are "essentially negligence questions."[388]

However, the return of negligence in design and warning defect cases can be overstated. First, the "mere supplier" or retailer may still be held liable for design or warning defects regardless of his or her level of care.[389] The first case extending strict liability to retailers was decided in 1964 in California,[390] and the reasons for the extension were instrumentalist.[391] This

---

(Ohio 1981) (citing *Greenman* and stating that "the policy underlying the doctrine" of strict liability was "that the public interest in human life and safety can best be protected by subjecting manufacturers of defective products to strict liability in tort when the products cause harm.").

380. *See* ABRAHAM, *supra* note 141, at 146.

381. *Id.*

382. *Id.*

383. *Id.*

384. *Id.* at 147.

385. *Id.*

386. *Id.*

387. *Id.*

388. *Id.*

389. *See* RESTATEMENT (SECOND) OF TORTS § 402A(2)(a) (1965).

390. *See* Vandermark v. Ford Motor Co., 391 P.2d 168 (Cal. 1964).

391. The California Supreme Court noted that retailers were "an integral part" of the distribution system and "should bear the cost of injuries resulting from defective products." *Id.* at 171. Then the court offered the familiar compensation and deterrence rationales. In certain cases the retailer "may be the only member of that enterprise reasonably available to the injured plaintiff." *Id.* Yet in other cases, "the retailer himself may play a substantial part in insuring that the product is safe or may be in a position to exert pressure on the manufacturer to that end; the retailer's strict liability thus serves as an added incentive to

holding has been widely followed in other jurisdictions[392] and makes the retailer's fault irrelevant.[393] Second, as Abraham noted, there are design and warning defect cases "that pay only lip service to these negligence-based standards when, in fact, loss-spreading or the insurance rationale appears to be the true basis for imposing liability."[394]

The instrumentalist rationales of compensation and deterrence were explicitly invoked, within tort itself, in the strict liability reform of products cases. The strictness of liability has been reduced in certain types of products cases by a return to something like negligence for manufacturers.[395] In any event, the portion of overall torts cases affected by instrumentalism in the reform of products liability is limited. It was not until the final category of tort reforms that instrumentalism achieved widespread effects.

4. "Modern" Tort Reforms

"Modern" tort reforms differ from the preceding reforms in several respects. First, the emphasis in the most recent reforms has not been on the standard of liability, as it was in workers' compensation, no-fault automobile laws, and products liability reform.[396] Instead, the primary focus has been on more peripheral issues, particularly damages.[397] Second, modern reforms have been cyclical, while the preceding reforms, with the possible exception of automobile accidents, occurred during a particular historical period.[398] Third, and most significant for present purposes, the modern reforms are generally undertaken for administrative efficiency.[399] Although workers' compensation and no-fault automobile laws were passed

---

safety." *Id.* at 171–72.

392. *See, e.g.*, Mead v. Warner Pruyn Div., Finch Pruyn Sales, Inc., 386 N.Y.S.2d 342 (N.Y. 1976), *aff'd*, 394 N.Y.S.2d 483 (N.Y. 1977); Seattle-First Nat'l Bank v. Volkswagen of Am., Inc., 525 P.2d 286 (Wash. Ct. App. 1974).

393. *See, e.g.*, Oser v. Wal-Mart Stores, 951 F.Supp. 115, 118–19 (S.D. Tex. 1996) ("[R]etailers as well as manufacturers can be liable for injuries caused by manufacturing and design defects . . . . Thus it does not matter whether [the retailer] can in any way be considered at fault for [a product's] potentially defective manufacture or design.").

394. ABRAHAM, *supra* note 141, at 148 (citing James A. Henderson Jr., *Echoes of Enterprise Liability in Product Design and Marketing Litigation*, 87 CORNELL L. REV. 958 (2002)).

395. Though it is beyond this article's scope, I do not concede that negligence is not, at least in part, based in instrumentalism. *See, e.g.*, Robert L. Rabin, *Some Thoughts on the Ideology of Enterprise Liability*, 55 MD. L. REV. 1190, 1199–1203 (1996); *see also supra* note 206.

396. *See* FRANKLIN, RABIN & GREEN, TORT LAW AND ALTERNATIVES 807 (8th ed., 2006).

397. *See id.*

398. *See id.* at 809–14 (discussing three "waves" of modern tort reform).

399. *See id.* at 807–08 (mentioning tort-related administrative costs).

for instrumentalist compensatory purposes and although products liability reform occurred for instrumentalist compensatory and deterrence purposes, modern tort reforms tend to be enacted to make the tort system, and enterprises it affects, perform more efficiently.[400]

Many of these reforms have been inspired by the cycle of liability insurance. Professor Kenneth Abraham described liability insurance as being subject to "periodic swings in the price and availability of coverage."[401] When the market is in a "soft" phase, premiums are relatively low and coverage is readily available.[402] However, when the "hard" phase arrives, premiums escalate quickly and precipitously and coverage can be difficult to obtain.[403] In the last forty years, the cycle has shifted into the hard phase three times: around 1975, around 1985, and in the early 2000s.[404] When the liability insurance cycle has become hard, it has been characterized as an insurance "crisis."[405] The premium increases and lack of insurance coverage have inspired legislatures to enact tort reforms for the sake of administrative efficiency.[406]

The hard cycles of 1975 and the early 2000s were largely limited to medical malpractice insurance.[407] Prodded by medical societies and the insurance industry, many state legislatures called for limits on damages, revocation of the collateral source rule, and generally making it harder to file medical malpractice suits, among other reforms.[408] One of the first, and certainly one of the most notable, tort reforms is California's Medical Injury Compensation Reform Act (MICRA), which was enacted in 1975.[409] Among other provisions, MICRA set a $250,000 cap on noneconomic damages for medical malpractice cases,[410] allowed periodic payments for "future damages" of $50,000 or more,[411] allowed disclosure of collateral

400. Whether these reforms are necessary or effective is beyond this article's scope. My sole claim is that these reforms are a pervasive source of instrumentalism in tort law.

401. ABRAHAM, *supra* note 141, at 121; *see also* FRANKLIN, RABIN & GREEN, *supra* note 396, at 809–14 (discussing liability insurance cycles).

402. *See* ABRAHAM, *supra* note 141, at 121.

403. *Id.*

404. *Id.*

405. *Id.*

406. *See infra* notes 407–45 and accompanying text.

407. ABRAHAM, *supra* note 141, at 121.

408. *See* FRANKLIN, RABIN & GREEN, *supra* note 396, at 809 ("Among the common changes in malpractice cases were placing caps on the amount that could be awarded for pain and suffering; regulating fees of plaintiffs' attorneys; shortening statutes of limitation; requiring periodic payments as to future awards; and altering or eliminating the collateral source rule.").

409. Medical Injury Compensation Reform Act of 1975, 1975 Cal. Stat. 3949 [hereinafter MICRA].

410. CAL. CIV. CODE § 3333.2 (West 1997).

411. CAL. CIV. PROC. CODE § 667.7 (West 2009).

sources of payment to the jury,[412] and set maximum percentages for contingency fees.[413] MICRA's preamble expressly provides the legislature's intent in its enactment:

> The Legislature finds and declares that there is a major health care crisis in the State of California attributable to skyrocketing malpractice premium costs and resulting in a potential breakdown of the health delivery system, severe hardships for the medically indigent, a denial of access for the economically marginal, and depletion of physicians such as to substantially worsen the quality of health care available to citizens of this state. The Legislature, acting within the scope of its police powers, finds the statutory remedy herein provided is intended to provide an adequate and reasonable remedy within the limits of what the foregoing public health and safety considerations permit now and into the foreseeable future.[414]

Plaintiffs and amici explicitly asked the California Supreme Court to re-investigate the "true" causes of the insurance crisis and to even "second-guess the [l]egislature as to whether a 'crisis' actually existed" at the time MICRA was passed.[415] The court declined, instead noting that the legislature had thoroughly investigated the matter and could rationally conclude "that the high insurance costs in this particular area posed special problems with respect to the continued availability of adequate insurance coverage and adequate medical care and could fashion remedies—directed to the medical malpractice context—to meet these problems."[416] Therefore, MICRA was intended, at least in part, to reduce the cost of medical malpractice liability insurance. In reducing these costs, the legislature wanted "(1) to restore insurance premiums to a level doctors and hospitals could afford, thereby inducing them to resume providing medical care to all segments of the community, and (2) to insure that insurance would in fact be available as a protection for patients injured through medical malpractice."[417] Thus, MICRA was enacted primarily for the instrumentalist goal of administrative efficiency. The aim was to lower insurance premiums so that doctors could afford liability insurance and continue to practice medicine in California. However, a secondary goal was also instrumentalist, and familiar: to provide compensation to those injured by malpractice.

In the wake of the second hard liability insurance cycle—largely limited to medical malpractice—many states passed tort reforms. For

---

412. CAL. CIV. CODE § 3333.1 (West 1997).

413. CAL. BUS. & PROF. CODE § 6146 (West 2003).

414. MICRA § 12.5, 1975 Cal. Stat. 3949, 4007.

415. Am. Bank & Trust Co. v. Cmty. Hosp. of Los Gatos-Saratoga, Inc., 683 P.2d 670, 678 (Cal. 1984).

416. *Id.*

417. *Id.*

example, Florida and Texas both passed tort reforms similar to California, and for similar reasons. In 2003, the Florida Legislature passed a noneconomic damages cap, which was set roughly at $500,000 against medical practitioners.[418] For emergency medicine practitioners, the noneconomic damages cap was set, in essence, at $150,000.[419] Just as for MICRA in California, the legislature focused on the cost of medical malpractice liability insurance; the "Legislative Findings and Intent" section provides:

> (a) Medical malpractice liability insurance premiums have increased dramatically in recent years, resulting in increased medical care costs for most patients and functional unavailability of malpractice insurance for some physicians.
> (b) The primary cause of increased medical malpractice liability insurance premiums has been the substantial increase in loss payments to claimants caused by tremendous increases in the amounts of paid claims.
> (c) The average cost of a medical negligence claim has escalated in the past decade to the point where it has become imperative to control such cost in the interests of the public need for quality medical services.
> (d) The high cost of medical negligence claims in the state can be substantially alleviated by requiring early determination of the merit of claims, by providing for early arbitration of claims, thereby reducing delay and attorney's fees, and by imposing reasonable limitations on damages, while preserving the right of either party to have its case heard by a jury.[420]

Just as in California, a court has examined the Florida law to determine its purpose.[421] A federal district court noted that both the statutory language cited above and a legislative summary provided to the governor after the bill passed both legislative chambers.[422] That document provided that all the evidence surveyed established that there was a medical malpractice insurance crisis "that threaten[ed] the quality and availability of health care for all Florida citizens."[423] The document stated, "Based on this record, this bill provides findings that making high quality health care available, ensuring physicians continue to practice, and ensuring the availability of affordable professional liability insurance to physicians are overwhelming public necessities."[424] The court specifically heralded the absence of a noneconomic damages cap as a cause of the insurance crisis.[425] Just as in

418. FLA. STAT. ANN. § 766.118(2) (West 2005 & Supp. 2010).
419. FLA. STAT. ANN. § 766.118(4) (West 2005 & Supp. 2010).
420. FLA. STAT. ANN. § 766.201 (West 2005).
421. Estate of McCall v. United States, 663 F. Supp. 2d 1276 (N.D. Fla. 2009).
422. *Id.* at 1299–1300.
423. *Id.* at 1300 (internal citation omitted).
424. *Id.*
425. *Id.*

California, plaintiffs challenged the findings, arguing that such findings were not entitled to deference.[426] The court rejected the plaintiff's argument, stating that the legislature had debated these issues and considered the evidence before making a rational policy choice.[427]

In 2003, Texas passed House Bill 4, which, among other provisions, limited recovery of noneconomic damages to $250,000 against all doctors and healthcare practitioners.[428] The legislature provided familiar reasoning for its passage.[429] In *Rivera v. United States*,[430] a federal district court

426. *Id.* at 1301.

427. *Id.*

428. TEX. CIV. PRAC. & REM. CODE ANN. § 74.301 (West 2005).

429. Section 10.11 provides:

> (a) The Legislature of the State of Texas finds that:
>
> (1) the number of health care liability claims (frequency) has increased since 1995 inordinately;
>
> (2) the filing of legitimate health care liability claims in Texas is a contributing factor affecting medical professional liability rates;
>
> (3) the amounts being paid out by insurers in judgments and settlements (severity) have likewise increased inordinately in the same short period;
>
> (4) the effect of the above has caused a serious public problem in availability of and affordability of adequate medical professional liability insurance;
>
> (5) the situation has created a medical malpractice insurance crisis in Texas;
>
> (6) this crisis has had a material adverse effect on the delivery of medical and health care in Texas, including significant reductions of availability of medical and health care services to the people of Texas and a likelihood of further reductions in the future;
>
> (7) the crisis has had a substantial impact on the physicians and hospitals of Texas and the cost to physicians and hospitals for adequate medical malpractice insurance has dramatically risen, with cost impact on patients and the public;
>
> (8) the direct cost of medical care to the patient and public of Texas has materially increased due to the rising cost of malpractice insurance protection for physicians and hospitals in Texas;
>
> (9) the crisis has increased the cost of medical care both directly through fees and indirectly through additional services provided for protection against future suits or claims, and defensive medicine has resulted in increasing cost to patients, private insurers, and Texas and has contributed to the general inflation that has marked health care in recent years;
>
> (10) satisfactory insurance coverage for adequate amounts of insurance in this area is often not available at any price;
>
> (11) the combined effect of the defects in the medical, insurance, and legal systems has caused a serious public problem both with respect to the availability of coverage and to the high rates being charged by insurers for medical professional liability insurance to some physicians, health care providers, and hospitals; and

interpreted the law's purposes and found that they "include reducing the frequency of healthcare liability claims, decreasing the amount of healthcare damages awards, protecting healthcare providers from potential liability, and making healthcare more accessible and affordable for Texas residents."[431]

In all three states—California in 1975 and Florida and Texas in the early-to-mid 2000s—the goal of tort reform was instrumentalist. Tort reform was the instrument that would lower medical malpractice liability insurance premiums. Lowering premiums would improve or retain the quality of healthcare by preventing medical providers from leaving the state or reducing services or by attracting new healthcare providers.[432] At least in California, MICRA satisfied a compensatory aim as well—the provision of a source of recovery (insurance) for injured patients.[433]

---

(12) the adoption of certain modifications in the medical, insurance, and legal systems, the total effect of which is currently undetermined, will have a positive effect on the rates charged by insurers for medical professional liability insurance.

(b) Because of the conditions stated in Subsection (a) of this section, it is the purpose of this article to improve and modify the system by which health care liability claims are determined in order to:

(1) reduce excessive frequency and severity of health care liability claims through reasonable improvements and modifications in the Texas insurance, tort, and medical practice systems;

(2) decrease the cost of those claims and ensure that awards are rationally related to actual damages;

(3) do so in a manner that will not unduly restrict a claimant's rights any more than necessary to deal with the crisis;

(4) make available to physicians, hospitals, and other health care providers protection against potential liability through the insurance mechanism at reasonably affordable rates;

(5) make affordable medical and health care more accessible and available to the citizens of Texas;

(6) make certain modifications in the medical, insurance, and legal systems in order to determine whether or not there will be an effect on rates charged by insurers for medical professional liability insurance; and

(7) make certain modifications to the liability laws as they relate to health care liability claims only and with an intention of the legislature to not extend or apply such modifications of liability laws to any other area of the Texas legal system or tort law.

H.B. 4, 2003 Leg., 78th Sess. (Tex. 2003), *available at* http://www.legis.state.tx.us/tlodocs/78R/billtext/pdf/HB00004F.pdf.

430. No. SA-05-CV-0101-WRF, 2007 WL 1113034 (W.D. Tex. Mar. 7, 2007).

431. *Id.* at *5.

432. *See supra* notes 409–31 and accompanying text.

433. *See supra* notes 409–17 and accompanying text.

However, medical malpractice is not the only tort doctrine affected by modern tort reforms that focus on administrative efficiency. The second "hard" cycle of liability insurance occurred in the mid-1980s and, unlike the first and third hard cycles, was not essentially restricted to medical malpractice liability insurance.[434] The tort reforms that were proposed, mostly in 1986 and 1987, focused on tort doctrines beyond medical malpractice.[435]

For example, California altered its doctrine of joint and several liability in 1986.[436] This common law doctrine allowed a plaintiff to recover the entire judgment from any of a number of defendants liable to the plaintiff for the same harm, as if that defendant were the sole tortfeasor.[437] On June 3, 1986, Californians passed Proposition 51, which effectively barred joint and several liability with respect to noneconomic damages.[438] One significant purpose behind its passage was administrative efficiency; specifically, the "Findings and Declaration of Purpose" references "soaring costs of lawsuits and insurance premiums," especially for local governments and concludes that reform is "necessary and proper to avoid catastrophic economic consequences."[439] The California Supreme Court

---

434. *See* ABRAHAM, *supra* note 141, at 121.

435. *See infra* notes 436–55 and accompanying text.

436. *See infra* note 438 and accompanying text.

437. *See, e.g.*, KENNETH S. ABRAHAM, THE FORMS AND FUNCTIONS OF TORT LAW 113 (3d ed. 2007); MARK A. GEISTFELD, TORT LAW: THE ESSENTIALS 404 (2008).

438. CAL. CIV. CODE § 1431.2 (West 2007); *see infra* note 440.

439. CAL. CIV. CODE § 1431.1 (West 2007). The entire section's language is:

> The People of the State of California find and declare as follows:
>
> (a) The legal doctrine of joint and several liability, also known as "the deep pocket rule", has resulted in a system of inequity and injustice that has threatened financial bankruptcy of local governments, other public agencies, private individuals and businesses and has resulted in higher prices for goods and services to the public and in higher taxes to the taxpayers.
>
> (b) Some governmental and private defendants are perceived to have substantial financial resources or insurance coverage and have thus been included in lawsuits even though there was little or no basis for finding them at fault. Under joint and several liability, if they are found to share even a fraction of the fault, they often are held financially liable for all the damage. The People—taxpayers and consumers alike—ultimately pay for these lawsuits in the form of higher taxes, higher prices and higher insurance premiums.
>
> (c) Local governments have been forced to curtail some essential police, fire and other protections because of the soaring costs of lawsuits and insurance premiums.

noted that the law, like many of its contemporaries, had been enacted due to "concerns over the effects of high liability insurance premiums."[440]

Also in 1986, the state of Washington enacted S.B. 4630, which included joint and several liability reform,[441] a variable cap on noneconomic damages,[442] and a provision for periodic payment of future damages.[443] Its preamble provides: "[I]t is the intent of the legislature to reduce costs associated with the tort system, while assuring that adequate and appropriate compensation for persons injured through the fault of others is available."[444] The legislature specifically mentioned that (1) "counties, cities, and other governmental entities are faced with increased exposure to lawsuits and awards and dramatic increases in the cost of insurance coverage," (2) there are "comparable cost increases in professional liability insurance" (including for physicians, architects, and engineers), and (3) "general liability insurance is becoming unavailable or unaffordable to many businesses, individuals, and nonprofit organizations in amounts sufficient to cover potential losses."[445] Moreover, "[h]igh premiums have discouraged socially and economically desirable activities and encourage many to go without adequate insurance coverage."[446]

That same year, Maryland enacted a $350,000 cap for noneconomic damages in personal injury actions generally.[447] The cap was enacted "in response to a legislatively perceived crisis concerning the availability and cost of liability insurance in this State."[448] The "crisis resulted in the unavailability of liability insurance for some individuals and entities, especially those engaged in hazardous activities such as asbestos removal,

---

> Therefore, the People of the State of California declare that to remedy these inequities, defendants in tort actions shall be held financially liable in closer proportion to their degree of fault. To treat them differently is unfair and inequitable.
>
> The People of the State of California further declare that reforms in the liability laws in tort actions are necessary and proper to avoid catastrophic economic consequences for state and local governmental bodies as well as private individuals and businesses.

Id.

440. Evangelatos v. Superior Court, 753 P.2d 585, 606 (Cal. 1988).

441. WASH. REV. CODE § 4.22.070(1)(b) (2008).

442. This provision was ruled unconstitutional under the state constitution in Sofie v. Fibreboard Corp., 771 P.2d 711, 728 (Wash. 1989).

443. WASH. REV. CODE § 4.56.260 (2008).

444. The preamble is contained in the "Historical and Statutory Notes" section of WASH. REV. CODE § 4.16.160 (2008).

445. *Id.*

446. *Id.*

447. MD. CODE ANN., CTS. & JUD. PROC. § 11-108(b) (LexisNexis 2006).

448. Murphy v. Edmonds, 601 A.2d 102, 114 (Md. 1992).

and increasing difficulty in obtaining reinsurance."[449] Thus, one of the cap's "primary purposes" was "to promote the availability and affordability of liability insurance in Maryland."[450] The cap was passed at the recommendation of two separate government task forces: the Governor's Task Force to Study Liability Insurance (issued December 20, 1985) and the Joint Executive/Legislative Task Force on Medical Malpractice Insurance (issued December 12, 1985).[451] The conclusion of the Governor's Task Force included the statement:

> The ceiling on noneconomic damages will help contain awards within realistic limits, reduce the exposure of defendants to unlimited damages for pain and suffering, lead to more settlements, and enable insurance carriers to set more accurate rates because of greater predictability of the size of judgments. The limitation is designed to lend greater stability to the insurance market and make it more attractive to underwriters.[452]

Modern tort reforms often pass for explicitly instrumentalist reasons, which include administrative efficiency. The goal of tort reform is to make the tort system, or activities that it affects, operate more smoothly. Chief among these concerns is the availability and affordability of liability insurance. Modern tort reforms have followed the insurance cycle; when insurance has become more expensive and more difficult to obtain, legislatures have acted. The reforms are not restricted to a particular type of tort law. Medical malpractice seems to be most affected by the cycle, but as the most recently discussed reforms indicate, these reforms are pervasive throughout tort law. Indeed, this discussion barely scratches the surface. One could continue and cover reforms wholly unrelated to the insurance cycle. For instance, some states have passed tort immunity statutes are designed to encourage volunteerism,[453] while other states have designed laws to protect recreational skiing for its positive economic effects,[454] and

---

449. *Id.* at 114–15.

450. Oaks v. Connors, 660 A.2d 423, 428 (Md. 1995).

451. *See* Franklin v. Mazda Motor Corp., 704 F. Supp. 1325, 1327 (D. Md. 1989).

452. *Id.* at 1328.

453. *See, e.g.*, COLO. REV. STAT. § 13-21-116(1) (2010) ("It is the intent of the general assembly to encourage the provision of services or assistance by persons on a voluntary basis to enhance the public safety rather than to allow judicial decisions to establish precedents which discourage such services or assistance to the detriment of public safety.").

454. *See, e.g.*, ALASKA STAT. § 5.45.010 (2008) ("Notwithstanding any other provision of law, a person may not bring an action against a ski area operator for an injury resulting from an inherent danger and risk of skiing.") The Legislative Findings and Purpose provide, in relevant part:

> (1) the sport of skiing is practiced by a large number of citizens of the state and also attracts a large number of nonresidents, providing significant contributions to the economy of the state through construction and operation

some states have immunized apologies by medical providers to encourage settlement of lawsuits.[455]

IV. CONCLUSION

Civil recourse theory is the most recent attempt to unify tort law. Writing from a positive perspective, Professors John C.P. Goldberg and Benjamin C. Zipursky have created a thoughtful and elegant explanation of tort law. Declaring tort the provision of recourse for those suffering deontic wrongs, Goldberg and Zipursky deny the existence of instrumentalism

---

> of skiing facilities, and through the money spent by citizens of the state and nonresidents;
> (2) the sport of skiing serves important public social and policy goals in the state given the dominance of the winter season; skiing contributes to the health and well-being of Alaskans, including the physically and mentally challenged; it is highly desirable that Alaskans have convenient and inexpensive access to the sport of skiing;
> . . . .
> (8) because of the size, power, and variation of the winter alpine environment, ski area operators are financially and physically incapable of controlling all the conditions under which skiing takes place.

ALASKA STAT., Temporary and Special Acts, Ch. 63 § 1 (1994).

455. *See, e.g.*, TENN. R. EVID. 409.1. The Advisory Commission Notes provide the purpose:

> Rule 409.1 renders inadmissible certain statements and actions reflecting sympathy for persons injured in accidents. This Rule, like Evidence Rules 408, 409, and 410, *is designed to encourage the settlement of lawsuits. It complements Evidence Rule 409, which makes inadmissible payment of medical and related expenses on the issue of liability. The underlying theory of Rule 409.1 is that a settlement of a lawsuit is more likely if the defendant is free to express sympathy for the plaintiff's injuries without making a statement that would be admissible as an admission of a party opponent.* Without this rule, a defendant's statement such is "I am sorry that you have suffered so much from the accident" might well be admissible as an admission of a party opponent. Accordingly, defense counsel may advise against making such statements in order to avoid the creation of harmful evidence. Yet a simple apology may go a long way toward making an injured party feel more comfortable with a nonjudicial settlement of the matter. This process is consistent with the modern focus on mediation and other methods of dispute resolution that seek to avoid a trial by facilitating a resolution acceptable to all parties.

Id. advisory comm'n cmt (emphasis added). Although this is technically an evidentiary doctrine, its aim is to affect tort cases.

within tort law. They do, however, acknowledge certain *de minimis* exceptions.

Yet the extent of instrumentalism in tort law is considerably greater than the scope of Goldberg and Zipursky's admissions. Using tort reform as a perspective by which to examine tort law, instrumentalism has grown increasingly pervasive through the progression of tort reforms over the last century. The first reform, workers' compensation, was based on the instrumentalist theory of compensation or loss spreading. However, it was implemented as a substitute for tort law and became very different from the tort system. The next major tort reform, no-fault for automobile accidents, was also based on instrumentalist compensation. By the time these reforms were debated and enacted, less of a sense of separation existed between tort law and the instrumentalist goal of compensation. Informally, tort law settlements had grown more oriented toward compensation. Furthermore, the formal reform doctrine was integrated into the existing tort system, as opposed to being independent thereof like workers' compensation.

The next major reform, products liability, was the first of the reforms explicitly based in instrumentalism, including both compensation and deterrence, and the first to be wholly within the law of torts. Yet the amount of cases that products liability reform actually affected was relatively small. Tort reform instrumentalism did not pervasively affect tort cases until the "modern" tort reforms. Based primarily on the instrumentalist goal of administrative efficiency, modern tort reforms have attempted to lower insurance premiums and tort judgments generally to improve the provision of healthcare and other services. Modern reforms affect tort doctrines across the spectrum, such as medical malpractice and products liability. Some states have even passed caps on noneconomic damages for all unintentional torts.

Thus, as a positive account of tort law, civil recourse theory is incomplete. An analysis of tort reform shows that instrumentalism, as a positive matter, is operating within tort law. Whether, as a normative matter, instrumentalism *should be* operating in tort law is a very different question. In her critique of civil recourse, Professor Stapleton, suggests that it would be more effective as a normative theory.[456] Whether pluralism is normatively and positively defensible is a project for another day.[457] But

---

456. Stapleton, *supra* note 16, at 1562.

457. Although a thorough discussion of that issue is beyond this article's scope, this analysis of tort reform offers a potential model for the relationship between "wrongs" and instrumentalism. A theme running through the reforms is the separation of personal injuries that should be approached instrumentally from those that should not.

Throughout the course of tort reforms, a consistent line of demarcation between treatment as a compensatory event and eligibility for "full-scale" tort treatment is wrongs (though perhaps not precisely as defined by Goldberg and Zipursky). The chief instrument for drawing this line is the general exclusion of intentional torts from instrumentalist limitations. Thus, employer-defendants committing intentional torts are generally not

even pluralism's most ardent defenders must acknowledge and respond to Goldberg and Zipursky's scholarship.

protected by workers' compensation exclusivity, drivers in no-fault jurisdictions generally are not protected from the full consequences of their intentional acts, and intentional torts are often excluded from noneconomic damages caps.

IN THE WAKE OF *THERASENSE & NISUS CORP.*, HOW CAN PATENT ATTORNEYS DEFEND *THEMSELVES* AGAINST ALLEGATIONS OF INEQUITABLE CONDUCT?

IAN G. MCFARLAND*

I. INTRODUCTION 488
II. PATENT LAW OVERVIEW 490
A. *Requirements and Procedures for Obtaining Letters Patent* 490
B. *Infringement* 492
C. *Inequitable Conduct* 493
III. EFFECTS OF AN INEQUITABLE CONDUCT FINDING ON A PATENT ATTORNEY 497
IV. POTENTIAL RECOURSE OPPORTUNITIES FOR PATENT ATTORNEYS FACING ALLEGATIONS OF INEQUITABLE CONDUCT 500
A. *Intervention* 500
1. Federal Rule of Civil Procedure 24 501
2. Standing 502
3. Intervention Requirements: Permissive vs. As-of-Right 506
4. Timeliness 510
5. *Nisus Corp.* and *Therasense* 513
B. *Appeal the District Court Opinion* 515
1. The "Nonparty Rule" and Types of Sanctions 516
2. Circuit Confusion: Appealing Sanctions 517
3. *Nisus Corp.* 521
C. *Petition for a Writ of Mandamus* 521
V. ANALYSIS 523
A. *Why the Federal Circuit Got Things Wrong in* Therasense 523
B. *Patent Attorneys Satisfy the Requirements for Permissive Intervention* 525
C. *A Statutory Right of Intervention is Necessary* 526
VI. CONCLUSION 529

* Candidate for Doctor of Jurisprudence, May 2011, University of Tennessee, College of Law. I would like to thank Professor Gary Puslinelli for his thoughtful insight and support.

"I STILL BELIEVE THAT MOST LAWYERS ARE WISE ENOUGH TO KNOW THAT THEIR MOST PRECIOUS ASSET IS THEIR PROFESSIONAL REPUTATION.
–JUSTICE JOHN PAUL STEVENS**

I. INTRODUCTION

By the summer of 2008, Lawrence Pope had been practicing patent law for over thirty-five years.[1] Attorney Pope was registered to practice in Illinois, Pennsylvania, the District of Columbia, and before the United States Patent and Trademark Office ("PTO").[2] For the better part of two decades, Attorney Pope had been employed by Abbott Laboratories ("Abbott"), where he prosecuted a significant number of their patents without ever being accused of inequitable conduct.[3] Needless to say, after thirty-five diligent and unblemished years, his professional reputation was certainly one of his most precious assets.

Meanwhile in the summer of 2008, Abbott's trial counsel was in its fourth year of fierce litigation in the United States District Court for the Northern District of California over the alleged infringement of four of Abbott's patents,[4] including U.S. Patent No. 5,820,551 ("the '551 patent"), a patent that was prosecuted by Lawrence Pope.[5] Becton, Dickinson & Company, Nova Biomedical Corporation, and Bayer Healthcare, LLC (collectively "BD/Nova") defended Therasense's claims of patent infringement with their own allegations that the '551 patent was unenforceable due to inequitable conduct on the part of Attorney Pope and Dr. Gordon Sanghera.[6]

On June 24, 2008, Lawrence Pope's spotless thirty-five year record of patent prosecution came to an end in a thirty-nine page opinion written by

---

** Cooter & Gell v. Hartmarx Corp., 496 U.S. 384, 413 (1990) (Stevens, J., concurring in part and dissenting in part).

1. Lawrence S. Pope's Motion For Leave To Intervene On Appeal at 1–2, Therasense, Inc. v. Becton, Dickinson & Co., 593 F.3d 1289 (Fed. Cir. 2010) (Nos. 2008-1511, -1512, -1513, -1514, -1595) [hereinafter "Pope's Motion to Intervene"], *available at* http://www.patentlyo.com/lawrence_20s._20pope_s_20motion_20for_20leave_20to_20intervene-1.pdf.

2. *Id.* at 2.

3. *Id.*

4. *See* Therasense, Inc. v. Becton, Dickinson & Co., 565 F. Supp. 2d 1088, 1091 (N.D. Cal. 2008).

5. Pope's Motion to Intervene, Declaration of Lawrence S. Pope at ¶ 5, Therasense, Inc. v. Becton, Dickinson & Co., 593 F.3d 1289 (Fed. Cir. 2010) (Nos. 2008-1511, -1512, -1513, -1514, -1595) [hereinafter Declaration of Lawrence S. Pope] ("I was responsible for the prosecution of the application that led to U.S. Patent No. 5,820,551 . . . ."). U.S. Patent '551 is entitled "Strip Electrode with Screen Printing" and "relates to enzymatic sensor electrodes and their combination with reference electrodes to detect a compound in a liquid mixture." U.S. Patent No. 5,820,551 col. 1 l. 25-27 (filed June 6, 1995).

6. *See Therasense*, 565 F. Supp. 2d at 1105.

Judge William Alsup.[7] Following the district court's decision, Abbott appealed the judgment to the Federal Circuit, preserving the issue of inequitable conduct.[8] Pope moved in the Federal Circuit for leave to intervene in the appeal,[9] a motion that was briskly denied in a one-page order.[10] Finally, on January 25, 2010, the Federal Circuit affirmed the district court's findings of inequitable conduct and judgment of unenforceability.[11]

Ultimately, Lawrence Pope stood defenseless as his thirty-five year record of credibility with the PTO dissolved during a legal proceeding to which he was not even a party. Following the district court's opinion, Attorney Pope received inquiries from both the Illinois Attorney Registration and Disciplinary Commission and the PTO's Office of Enrollment and Discipline.[12] Additionally, no sooner than the ink had dried on the district court's opinion, Pope's professional reputation was attacked by the press.[13]

A patent attorney[14] who has prosecuted a patent that later becomes the subject of inequitable conduct claims has limited recourse in defending his or her professional reputation in light of those allegations. In those

---

7. *See id.* at 1114 ("The withheld extrinsic evidence here was richly material. And, intent to deceive, not just to withhold, was clearly in the mind of Attorney Pope, hard as it is to so conclude as to a professional.").

8. *See* Therasense, Inc. v. Becton, Dickinson & Co., 593 F.3d 1289, 1292 (Fed. Cir. 2010).

9. Pope's Motion to Intervene, *supra* note 1, at 1.

10. Order Before Gajarsa, J., Therasense v. Becton, Dickinson & Co., 593 F.3d 1289 (Fed. Cir. 2008) (Nos. 2008-1511, -1512, -1513, -1514, -1595) [hereinafter *Therasense* Order], *available at* http://www.patentlyo.com/files/order-re-lawrence-s.-pope_s-motion-for-leave-to-intervene-in-appeal.pdf.

11. Therasense, Inc. v. Becton, Dickinson & Co., 593 F.3d 1289, 1307 (Fed. Cir. 2010) ("Because the district court's findings that the EPO submissions were highly material to the prosecution of the '551 patent and that Pope and Dr. Sanghera intended to deceive the PTO by withholding those submissions were not clearly erroneous, the district court did not abuse its discretion in holding the '551 patent unenforceable due to inequitable conduct."), *vacated*, 374 F. App'x 35 (Fed. Cir. 2010). The Federal Circuit has ordered a rehearing en banc and has designated six issues for review. *Therasense*, 374 F. App'x 35. However, the order denying Pope's Motion For Leave to Intervene is not one of those issues. *Id.*

12. Declaration of Lawrence S. Pope, *supra* note 5, at ¶ 7.

13. *See* Zusha Elinson, *Judge Kills IP Claim, Blames Lawyer*, RECORDER, June 26, 2008, *available at* http://www.law.com/jsp/article.jsp?id=1202422556384; *see also* Amy Miller with additional reporting by Zusha Elinson, *Preferable Isn't Required*, CORPORATE COUNSEL, at 17, Sept. 1, 2008, *available at* http://www.law.com/jsp/cc/PubArticleCC.jsp?id=1202423724483 (discussing how Abbott's patent was declared unenforceable "thanks to its lawyer's creative wordplay.").

14. In this Comment, "Patent attorney" can also include patent agents in certain contexts. Patent agents are individuals registered to practice before the PTO who are not attorneys. 37 C.F.R. § 11.6(b) (2009). "Patent practitioner" is also a term that can be used to encompass both practices. *See* 37 C.F.R. § 1.32(a) (2009).

situations, the patent attorney should be permitted to intervene in the proceedings as long as the original parties are still disputing that issue.

This Comment will provide a brief overview of the *Therasense* and *Nisus Corp.* cases and their application of inequitable conduct. It will analyze the practical effects on attorneys who have been the subject of inequitable conduct claims and findings. It will then discuss the possible recourse opportunities for patent attorneys who are faced with inequitable conduct claims at trial and evaluate their feasibility. Finally, this Comment will counsel for reform that would allow patent attorneys to intervene where prosecution of their patent involves the issue of inequitable conduct.

Part II of this Comment will provide a general overview of the patent law regime, including a discussion of both infringement and inequitable conduct. Part III will discuss the effects that inequitable conduct, and discipline in general, have on practicing patent attorneys.[15] Part IV will focus on the potential recourse opportunities that patent attorneys have (or do not have) to defend themselves in light of inequitable conduct claims, including intervention, appeals, and writs of mandamus. The practical feasibility of each of these opportunities is also evaluated, as well as the impact that these doctrines had on *Nisus Corp.* and *Therasense*. Part V will explain why the Federal Circuit was wrong in not allowing Attorney Pope to intervene in *Therasense*. Furthermore, it will show why patent attorneys satisfy the requirements for permissive intervention, and it counsels for a statutory right of intervention to defend against claims of inequitable conduct. Part VI will offer a brief conclusion.

II. PATENT LAW OVERVIEW

A. *Requirements and Procedures for Obtaining Letters Patent*

Article I of the United States Constitution authorizes Congress "[t]o promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries[.]"[16] Pursuant to this power, Congress passed the United States Patent Act of 1952, which has served as the foundation for our current patent regime since its inception.[17] The cornerstone of the Act reads: "[w]hoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, [sic] subject to the conditions and requirements of this title."[18]

---

15. Aside from formal discipline, the public scorn often does the most damage to the careers of patent attorneys.

16. U.S. CONST. art. I, § 8, cl. 8.

17. U.S. Patent Act of 1952, ch. 950, 66 Stat. 792 (codified as Title 35 of the United States Code, entitled "Patents").

18. 35 U.S.C. § 101 (2006); *see also id.* § 161 (discussing requirements for plant

The PTO is responsible for collecting patent applications, reviewing the applications, and issuing letters patent to the deserving applicants.[19] Patent applications contain a specification of the invention.[20] This includes a "written description" of the invention, the "claims" or limits of the invention, and its "best mode," at least one drawing of the invention, and an oath from each applicant that the applicant is the true inventor.[21] Patent applications are reviewed in light of the following requirements for obtaining a patent: patentable subject matter,[22] utility,[23] novelty,[24] non-obviousness,[25] and disclosure.[26]

Upon receiving a reviewable patent application, an examiner begins by searching for relevant "prior art" in the PTO's library or the examiner's own private resource materials.[27] After determining the results of the search, the examiner's position with regard to patentability is communicated to the applicant through what is known as an "Office Action."[28] Typically, patent applications are initially rejected through an Office Action for technical formalities, patentability, or both.[29]

Applicants may respond to Office Actions by amending their claims, by offering arguments in rebuttal to the Office Action, or a combination thereof.[30] Affidavits attesting to non-obviousness or the date of invention may also accompany a response by the applicant.[31] This exchange between

---

patents); *id.* § 171 (discussing requirements for design patents).

19. *See id.* § 2.

20. *Id.* §§ 111(a), 112, 113, 115.

21. According to §§ 111(a), 112, 113, 115:

> The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor of carrying out his invention.
>
> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

22. *Id.* at § 101 ("process, machine, manufacture, or composition of matter").

23. *Id.* ("useful").

24. *Id.* ("new"); *id.* § 102.

25. *Id.* § 103 ("A patent may not be obtained . . . if the differences between the subject matter sought to be patented and the prior art . . . would have been obvious at the time the invention was made to a person having ordinary skill in the art . . . .").

26. *Id.* § 112.

27. HERBERT F. SCHWARTZ, PATENT LAW & PRACTICE 14–15 (2d ed. 1995).

28. *Id.* at 15.

29. *Id.* at 15–16.

30. *Id.* at 16–17.

31. *Id.* at 17.

the examiner and the patent attorney, as well as the payment of regular fees,[32] comprises the process of prosecuting a patent.

However, one important caveat of this practice is that no new matter may be added to the application.[33] The applicant, via a patent attorney or agent, is allowed to amend their claims as much as the applicant wants but may add nothing to the originally filed specification.[34] Therefore, because of the disclosure requirement for patentability,[35] the applicant's ability to modify the claims is limited by what was initially disclosed in the written description.[36]

After a series of Office Actions and responses, applicants hope to receive a Notice of Allowance from the examiner, which is required for the inventor to obtain the letters patent.[37] Patents are valid for twenty years, beginning on the filing date of the application,[38] which can take over thirty-four months to prosecute.[39]

B. *Infringement*

The owner of a patent possesses an absolute right to exclude others from making, using, selling, offering to sell, and importing the patented invention in the United States for the duration of the patent.[40] A patent owner's invention is described by the claims in the patent application.[41] Therefore, an allegedly infringing item is compared to the claim language, and if a corresponding element on the item matches with every limitation of the claim, the item infringes the patent.[42]

A variety of legal remedies are available to the owner of an infringed patent. Preliminary and permanent injunctions[43] are frequently issued as well as monetary damages for lost profits,[44] reasonable royalties,[45] attorneys fees,[46] and trebled damages.[47] Infringement defense can, for the most part,

---

32. *See id.* at 13 ("Additional fees may be incurred during prosecution of the application if the number and form of the claims are changed by amendment.").

33. *See* 35 U.S.C. § 132 (2006).

34. *See id.*; SCHWARTZ, *supra* note 27, at 17.

35. 35 U.S.C. § 112.

36. *See* Gentry Gallery, Inc. v. Berkline Corp., 134 F.3d 1473, 1479 (Fed. Cir. 1998) ("[T]he scope of the right to exclude may be limited by a narrow disclosure.").

37. *See* SCHWARTZ, *supra* note 27, at 17.

38. 35 U.S.C. § 154(a)(2).

39. *Patent Pendency Statistics – FY09*, U.S. PATENT & TRADEMARK OFFICE, http://www.uspto.gov/patents/stats/patentpendency.jsp (last modified Oct. 7, 2009).

40. 35 U.S.C. § 271(a).

41. SCHWARTZ, *supra* note 27, at 73.

42. *See id.* at 79–81.

43. 35 U.S.C. § 283.

44. *Id.* § 284.

45. *Id.*

46. *Id.* § 285.

be characterized as a two-front battle. Accused infringers defend their items against the language of the patent claim by arguing non-infringement because the claim's limitations do not all have corresponding elements on their item.[48] Defendants also attack the patent on validity grounds by arguing that the patent should never have been issued based on one or more requirements for patentability.[49] Similar to invalidity, allegations that the patent is unenforceable, typically due to inequitable conduct before the PTO, are also an option.[50]

C. Inequitable Conduct

Inequitable conduct is an equitable defense used by accused infringers to avoid liability altogether.[51] Since the United States Supreme Court initially endorsed the doctrine in 1945,[52] defendants have attempted to use it in the vast majority of patent infringement lawsuits.[53] Many commentators, and even the Federal Circuit, have described this trend as an "absolute plague."[54] In fact, Chief Judge Rader has gone so far as to characterize this doctrine as that "which grew from a tiny bush on the patent landscape that inhibited gross fraud into a ubiquitous weed that infects every prosecution and litigation involving patents."[55]

Inequitable conduct occurs when a patent applicant affirmatively misrepresents material facts to the examiner, submits false information, or fails to disclose material information such as relevant prior art.[56] Additionally, such misrepresentations or omissions must bear evidence of intent to deceive.[57] Materiality and intent must both be proven by clear and convincing evidence and then balanced against one another on a sliding scale, "with a greater showing of one factor allowing a lesser showing of the other."[58] The Federal Circuit reviews district court findings of inequitable conduct for abuse of discretion.[59]

---

47. *Id.* § 284.

48. SCHWARTZ, *supra* note 27, at 88.

49. *Id.*

50. *See id.* at 89–92.

51. *See id.*

52. Precision Instrument Mfg. Co. v. Auto. Maint. Mach. Co., 324 U.S. 806, 817–20 (1945) (citing Hazel-Atlas Co. v. Hartford Co., 322 U.S. 238, 246 (1944) (dictum)).

53. SCHWARTZ, *supra* note 27, at 90.

54. Burlington Indus., Inc. v. Davco Corp., 849 F.2d 1418, 1422 (Fed. Cir. 1988).

55. The Honorable Randall R. Rader, *Always at the Margin: Inequitable Conduct in Flux*, 59 AM. U. L. REV. 777, 781 (2010).

56. Molins PLC v. Textron, Inc., 48 F.3d 1172, 1178 (Fed. Cir. 1995).

57. *Id.*

58. Union Pac. Res. Co. v. Chesapeake Energy Corp., 236 F.3d 684, 693 (Fed. Cir. 2001).

59. Kingsdown Med. Consultants, Ltd. v. Hollister Inc., 863 F.2d 867, 876 (Fed. Cir. 1988) (en banc).

PTO Rule 56 mandates that:

> Each individual associated with the filing and prosecution of a patent application has a duty of candor and good faith in dealing with the Office, which includes a duty to disclose to the Office all information known to that individual to be material to patentability as defined in this section.[60]

The Rule further states that the materiality requirement is met if the information establishes or contributes to a prima facie case of unpatentability or if it contradicts an applicant's position when "[o]pposing an argument of unpatentability relied on by the Office, or . . . [a]sserting an argument of patentability."[61]

The Federal Circuit has insisted that "materiality does not presume intent. . . ."[62] A threshold level of intent must be determined independently from materiality.[63] Moreover, "specific intent to . . . mislead[ ] or deceiv[e] the PTO" must be proven.[64] In *Molins PLC v. Textron, Inc.*, the Federal Circuit characterized this requirement of specificity as the applicant making "a *deliberate decision* to withhold a *known* material reference."[65]

Two important cases highlight this application of the doctrine from the point of view of the accused patent attorney. In the first case, *Nisus Corp. v. Perma-Chink Systems, Inc.*, the accused infringer alleged inequitable conduct based on a failure to disclose to the PTO documents material to the patent as well as documents that came to light in an earlier lawsuit involving similar patents.[66] Patent attorney Michael Teschner began the prosecution of Nisus's patent application that eventually became United States Patent No. 6,426,095 B2 ("the '095 patent").[67] At some point during the prosecution, just prior to the deadline to submit pertinent materials to the PTO examiner, Teschner was replaced by another patent attorney named Allan Altera.[68] Attorney Teschner insisted that he turned over all materials relevant to the prosecution of the '095 patent to Altera during the transition, and that this should have fulfilled his duty of disclosure because the deadline for submitting materials had not yet lapsed.[69] Nonetheless, the district court was unconvinced and still specifically characterized Attorney

---

60. 37 C.F.R. § 1.56(a) (2009).

61. *Id.* at § 1.56(b)(2).

62. Allen Organ Co. v. Kimball Int'l, Inc., 839 F.2d 1556, 1567 (Fed. Cir. 1988).

63. *See id.*

64. Star Scientific, Inc. v. R.J. Reynolds Tobacco Co., 537 F.3d 1357, 1366 (Fed. Cir. 2008) (quoting Molins PLC v. Textron, Inc., 48 F.3d 1172, 1181 (Fed. Cir. 1995)) (alteration in original) (emphasis omitted).

65. *Molins*, 48 F.3d at 1181 (emphasis added).

66. Nisus Corp. v. Perma-Chink Sys., Inc., 421 F. Supp. 2d 1084, 1088 (E.D. Tenn. 2006).

67. *Id.* at 1096–97.

68. *Id.* at 1097–98.

69. *Id.* at 1097.

Teschner's conduct as inequitable.[70] The district court concluded that Attorney Teschner's failure to disclose to the PTO documents from a related litigation[71] during his tenure as prosecuting counsel of the '095 patent was inexcusable.[72] According to the court, Teschner "acknowledge[d] that he intentionally did not disclose Nisus I . . . ."[73] Furthermore, documents that were not disclosed were ones "that the PTO has specifically required applicants to disclose—namely, evidence of prior public use or sale and evidence inconsistent with Nisus's assertions regarding patentability."[74] Finally, the court characterized Attorney Teschner's attempted explanation of his conduct as not credible.[75]

The second case, *Therasense v. Becton, Dickinson & Co.*, involved U.S. Patent No. 5,820,551 ("the '551 patent"), which taught a disposable, electrochemical diabetic blood testing kit with a sensor that lacked a membrane.[76] The prior art electrochemical sensors contained membranes to prevent red blood cells from sticking to the electrode and interfering with the results.[77] However, one of Therasense's prior patents, U.S. Patent Nos. 4,545,382 ("the '382 patent"), disclosed an *optional*-membrane electrochemical sensor.[78] Because the '382 patent was filed well before the '551 patent, Therasense repeatedly faced rejection of the '551 patent over the '382 patent for anticipation and obviousness.[79] Indeed, from 1984, when the application that matured into the '551 patent was filed, until 1997, Therasense pursued the patent's prosecution through several continuation applications.[80]

At issue in the prosecution of the '551 patent was the '382 patent's "optionally, but preferably" language when describing the need for a protective membrane to be used on whole blood.[81] The examiner initially understood this language to mean that a membrane was not required for the '382 patent either, rendering the '551 patent obvious and anticipated.[82] However, Therasense's patent attorney, Lawrence Pope, insisted that when

---

70. *See id.* at 1106–07.

71. *See* Nisus Corp. v. Perma-Chink Sys., Inc., 327 F. Supp. 2d 844 (E.D. Tenn. 2003) ("Nisus I").

72. *See Nisus Corp.*, 421 F. Supp. 2d at 1102–03 ("During the six-month period Mr. Teschner bore the duty of disclosure on behalf of Nisus, there is no written record that he disclosed Nisus I or any Nisus I-related documents to the Patent Examiner.").

73. *Id.* at 1103.

74. *Id.* at 1104.

75. *Id.* at 1106.

76. Therasense, Inc. v. Becton, Dickinson & Co., 593 F.3d 1289, 1293–94 (Fed. Cir. 2010).

77. *Id.* at 1294.

78. *Id.*

79. *Id.* at 1301.

80. *Id.*

81. *Id.* at 1294, 1302 (citation omitted).

82. *Id.* at 1301.

the '551 patent was filed, one skilled in the art would have understood the language to mean that a membrane was required for whole blood in the '382 patent.[83] Attorney Pope further represented to the PTO that the "optionally, but preferably" language was "mere patent phraseology."[84] After Attorney Pope filed a declaratory affidavit to cement the argument, the PTO allowed the '551 patent with a claim for a membraneless sensor.[85]

The precise finding of inequitable conduct resulted from the court's hearing evidence regarding a representation made to the European Patent Office ("EPO") when prosecuting the '382 patent's European counterpart—European Patent EP 0 078 636 ("the '636 patent")—which was not disclosed to the examiner during the U.S. prosecution.[86] In 1993, the EPO revoked the '636 patent for obviousness over a German, prior art reference labeled D1, which described a sensor with a semi-permeable membrane.[87] The '636 patent contained the same "optionally, but preferably" language as the '382 patent.[88] To defeat the EPO's attack on the '636 patent, Therasense filed a brief and maintained that the "optionally, but preferably" language was "unequivocally clear. The protective membrane is optional . . . ."[89]

The district court concluded that the representations made to the EPO about the '382 patent's European counterpart were highly material and inconsistent with the arguments made to the PTO concerning the '551 patent.[90] The court had a particular problem with the fact that the EPO was told that the '636 patent's "optionally, but preferably" language was "unequivocally clear," but that Therasense insisted to the PTO that the language was "mere patent phraseology."[91] The district court also concluded that Attorney Pope "made a conscious decision to withhold the contradictory material from the PTO."[92] The entire '551 patent was held unenforceable because of inequitable conduct.[93]

---

83. *Id.* at 1301–02.

84. *Id.* at 1302.

85. *Id.*

86. *Id.*

87. *Id.*

88. *Id.*

89. *Id.* at 1303.

90. Therasense, Inc. v. Becton, Dickinson & Co., 565 F. Supp. 2d 1088, 1112 (N.D. Cal. 2008).

91. *Id.* at 1107–08.

92. *Id.* at 1113.

93. *Id.* at 1127.

III. EFFECTS OF AN INEQUITABLE CONDUCT FINDING ON A PATENT ATTORNEY

According to Judge Rader, "An allegation of inequitable conduct open[s] the door to vast discovery into the circumstances of the patent prosecution, level[s] an embarrassing charge of fraud as a counterweight to the presumption of patent validity, and even disqualifie[s] the prosecuting attorney (who may be a witness) from the patentee's litigation team."[94] Without the ability to use any of the recourse opportunities discussed in Part IV, patent attorneys will have no choice but to accept the district court's findings and deal with the consequences.[95] Aside from the public ridicule of the district court opinion, patent attorneys can face discipline from the PTO.[96] Congress has authorized the Commissioner of the PTO to regulate and discipline patent attorneys and agents.[97] In fact, twenty-eight final orders of discipline have been issued by the PTO's Office of Enrollment and Discipline ("OED") in the past year.[98] Furthermore, pursuant to the Freedom of Information Act ("FOIA"), these records are accessible to the public.[99]

Regulations direct the Commissioner of the PTO to appoint a Committee on Discipline, consisting of "at least three employees of the Office."[100] The Committee's primary responsibility is to determine and recommend whether there is sufficient probable cause against a practitioner to bring a charge.[101] Prior to determining whether sufficient probable cause exists, the OED investigates "information or evidence from any source suggesting possible grounds for discipline."[102] In order to proceed in a disciplinary action against a practitioner, the OED must prove a violation by clear and convincing evidence.[103] Furthermore, a five-year statute of limitations forces the OED to conduct investigations in a timely, efficient manner.[104]

Inequitable conduct by a patent attorney clearly violates the PTO's Disciplinary Rules.[105] "Knowingly giving false or misleading information

94. Rader, *supra* note 55, at 783.

95. *See infra* Part IV.

96. *See* 35 U.S.C. § 32 (2006).

97. *Id.*

98. *See Decisions of the Office of Enrollment and Discipline*, UNITED STATES PATENT AND TRADEMARK OFFICE, http://des.uspto.gov/Foia/OEDReadingRoom.jsp (search with "Decision Type" of "Discipline," "Start Date" of "06/01/2009," and "End Date" of "06/01/2010").

99. *See* 5 U.S.C. § 552 (2006).

100. 37 C.F.R. § 11.23(a) (2009).

101. *Id.* § 11.23(b).

102. *Id.* § 11.22(a).

103. *Id.* § 11.49.

104. 28 U.S.C. § 2462 (2006).

105. *See* 37 C.F.R. § 10.23(c)(2)(ii) (2009); 37 C.F.R. § 11.19(b)(iv).

or knowingly participating in a material way in giving false or misleading information, to . . . the Office or any employee of the Office"[106] constitutes grounds for discipline.[107] Not surprisingly, this is the same standard that governs the inequitable conduct doctrine in the courts.[108] Moreover, because an accused infringer must prove inequitable conduct by clear and convincing evidence,[109] any district court opinion containing such a finding will have already satisfied the OED's standard.

Once a charge is brought by the OED, the attorney may answer the complaint and contest the charges in front of a hearing officer.[110] The respondent to a charge has the burden of proving any affirmative defense by clear and convincing evidence.[111] After listening to both sides' arguments, the hearing officer renders an initial decision,[112] which either party may appeal to the Director of the USPTO.[113]

Following the hearings and appeals, the PTO has four types of discipline available to it if it finds a violation: exclusion, suspension, reprimand or censure, and probation.[114] More likely, however, is that a patent attorney will resign from the PTO prior to the issuance of any complaint or decision, avoiding public discipline.[115] For example, in *Molins PLC v. Textron, Inc.*, the district court found that the patent attorney's failure to cite a material set of articles during the patent's prosecution constituted inequitable conduct, and the patent was declared unenforceable.[116] Following the district court's decision, the patent attorney resigned from the patent bar.[117]

In *Applied Materials, Inc. v. Multimetrixs*, the district court found a patent unenforceable due to inequitable conduct when one of the inventors' signatures on a disclosure made to the PTO was forged.[118] One of the three inventors died during the prosecution of the patent.[119] Although, the patent

---

106. *Id.* § 10.23(c)(2)(ii).

107. *Id.* § 11.19(b)(iv).

108. Therasense, Inc. v. Becton, Dickinson & Co., 593 F.3d 1289, 1300 (Fed. Cir. 2010) (citing Innogenetics, N.V. v. Abbott Labs., 512 F.3d 1363, 1378 (Fed. Cir. 2008)).

109. Kingsdown Med. Consultants, Ltd. v. Hollister Inc., 863 F.2d 867, 876 (Fed. Cir. 1988) (en banc).

110. 37 C.F.R. §§ 11.36–39 (2009).

111. *Id.* § 11.49.

112. *Id.* § 11.54.

113. *Id.* § 11.55.

114. *Id.* § 11.20.

115. *See, e.g.*, Molins PLC v. Textron, Inc., 48 F.3d 1172 (Fed. Cir. 1995); Edwin S. Flores & Sanford E. Warren, Jr., *Inequitable Conduct, Fraud, and Your License to Practice Before the United States Patent and Trademark Office*, 8 TEX. INTELL. PROP. L.J. 299, 322 (2000).

116. *Molins*, 48 F.3d. at 1177.

117. *See* Flores & Warren, *supra* note 115, at 322.

118. No. C 06-07372 MHP, 2008 WL 2892453, at *4–9 (N.D. Cal. 2008).

119. *Id.* at *2.

attorney did not learn of the inventor's death until years later, the OED found that the attorney had committed inequitable conduct sufficient to warrant discipline because he failed to fulfill his duty "to conduct an inquiry reasonable under the circumstances."[120] The two remaining inventors filed a document with the patent attorney that only they had signed.[121] After the PTO rejected the document because all the inventors had not signed it, the document was remedied to bear all three signatures, not by the patent attorney, but by the inventors.[122] Even though the patent attorney had not known what had occurred, the OED felt that the remedy should have triggered an inquiry on the part of the patent attorney.[123] The OED publicly reprimanded the patent attorney.[124]

Not only do patent attorneys who were the subject of inequitable conduct findings have to worry about the threat of formal discipline from the PTO, but they must also deal with the scorn of public ridicule.[125] For instance, after the district court's opinion in *Therasense* was published, two separate magazine articles commented specifically on Attorney Pope's conduct.[126] The consequences that publication in print media will have on the patent attorney are impossible to measure. However, clients and prospective clients reading the article undoubtedly may decide to take their business elsewhere. Indeed, an attorney's record relating to inequitable conduct is extremely relevant information when a firm decides to hire counsel for prosecution. In an effort to attain the strongest patent possible and thereby dissuade any future infringers from challenging the patent in court, firms will likely gravitate toward attorneys who have not been the subject of inequitable conduct findings.

Attorney Pope has faced both disciplinary consequences from the PTO and public ridicule in the press. According to his declaratory affidavit, following the court's ruling that he had committed inequitable conduct, he "has received inquiries from both the Illinois Attorney Registration and Disciplinary Commission and the PTO's Office of Enrollment and Discipline."[127] Moreover, in one article, a commentator points out that even if the OED is precluded by the statute of limitations from imposing discipline on Pope as a result of the *Therasense* findings, "[t]hat doesn't mean that [the office] can't look at other cases Pope has executed . . . just to see if there are any other inequitable conduct cases that may have arisen."[128]

---

120. *In re* Zborovsky, Final Order, USPTO Proc. No. D09-34 at 5 (Aug. 31, 2009).
121. *Applied Materials*, 2008 WL 2892453 at *3.
122. *Id.* at *3.
123. *See Zborovsky*, Final Order at 5.
124. *See id.* at 9.
125. *See, e.g.*, Elinson, *supra* note 13; Miller, *supra* note 13, at 17–18.
126. *See* Elinson, *supra* note 13; Miller, *supra* note 13, at 17–18.
127. Pope's Motion to Intervene, *supra* note 1, at 1.
128. Miller, *supra* note 13, at 18.

Therefore, it appears that the validity of Pope's entire career is now in question after one finding of inequitable conduct by the district court.

IV. POTENTIAL RECOURSE OPPORTUNITIES FOR PATENT ATTORNEYS FACING ALLEGATIONS OF INEQUITABLE CONDUCT

A. Intervention

The most direct and effective way for patent attorneys to defend themselves against inequitable conduct charges would be to intervene in the district court proceedings. After all, the most critical occurrence for patent attorneys is when the district court enters its opinion on the issue, crystallizing the finding of inequitable conduct.[129] Intervening as a party prior to the case's disposition allows individuals to present their own evidence and arguments on the issue in hopes of avoiding the scathing commentary associated with findings of inequitable conduct.[130]

Traditionally, three judicial doctrines have stood in the way of motions for intervention.[131] Potential intervenors must overcome issues of standing, timing, and the adequacy of their interests.[132] The standards by which these issues are analyzed, however, vary depending upon whether intervention is sought permissively or as of right.[133] An exhaustive discussion of the complexities presented by issues of intervention, or a complete dissection of the differences between the two intervention provisions, is beyond the scope of this Comment. Rather, the goal of this section is to highlight and analyze the potential implications that these issues and differences have for the ability of patent attorneys to intervene in suits involving inequitable conduct charges against them.

Initially, this section will introduce Federal Rule of Civil Procedure 24 ("Rule 24") and emphasize some similarities and differences between intervention as of right and permissive intervention.[134] A discussion of the standing doctrine and its effects on intervention will follow as well as an examination of the differing approaches that the Circuit Courts of Appeals have developed regarding the relationship between these two doctrines as a result of limited guidance from the Supreme Court.[135] Next, the various elements of both intervention provisions in Rule 24 will be examined,[136] followed by a discussion on timeliness that includes the proposition of

129. *See* Int'l Union v. Scofield (*Auto Workers*), 382 U.S. 205, 213 (1965).
130. *See* FED. R. CIV. P. 24 (permitting parties to intervene in district court).
131. *See infra* Section IV(A)(2)–(4).
132. *See id.*
133. *See infra* Section IV(A)(1) & (3).
134. *See infra* Section IV(A)(1).
135. *See infra* Section IV(A)(2).
136. *See infra* Section IV(A)(3).

intervening in an appeal.[137] Finally, this section will conclude with an analysis of the role intervention played in both *Nisus Corp.* and *Therasense.*[138]

1. Federal Rule of Civil Procedure 24

Although intervention as of right through Rule 24(a) and permissive intervention through Rule 24(b) are completely separate avenues, it is hard to ignore some of their obvious similarities.[139] First, both provisions allow for statutory intervention in some capacity.[140] While Rule 24(a) speaks of unconditional statutory rights, thus *requiring* intervention, Rule 24(b) addresses conditional statutory rights that leave room for judicial discretion.[141] Furthermore, both provisions require that the applicant's motion be timely.[142]

The most significant differences between the two provisions are the way in which the potential intervenor's interests are measured.[143] Intervention as of right requires "an interest relating to the property or transaction that is the subject of the action."[144] Additionally, Rule 24(a) mandates that the potential intervenor be in a position such "that disposing of the action may as a practical matter impair or impede the movant's ability to protect its interest."[145] Finally, existing parties already in the litigation must not already "adequately represent that interest."[146]

These requirements stand in contrast to those of Rule 24(b), that the potential intervenor "has a claim or defense that shares with the main action a common question of law or fact."[147] Along with sharing common factual or legal questions, potential intervenors must also convince the courts that allowing them to intervene will not "unduly delay or prejudice the adjudication of the original parties' rights."[148] Clearly, however, the requirements for permissive intervention are easier to satisfy than those of its counterpart.

---

137. *See infra* Section IV(A)(4).

138. *See infra* Section IV(A)(5).

139. *See* FED. R. CIV. P. 24(a), (b).

140. FED. R. CIV. P. 24(a)(1) & (b)(1)(A).

141. *Compare* FED. R. CIV. P. 24(a) ("On timely motion, the court *must* permit anyone to intervene who . . . .") (emphasis added), *with* FED. R. CIV. P. 24(b)(1) ("On timely motion, the court *may* permit anyone to intervene who . . . .") (emphasis added).

142. FED. R. CIV. P. 24(a) & (b).

143. *See* FED. R. CIV. P. 24(a)(2); FED. R. CIV. P. 34(b)(1)(B); FED. R. CIV. P. 24(b)(3).

144. FED. R. CIV. P. 24(a)(2).

145. *Id.*

146. *Id.*

147. FED. R. Civ. P. 24(b)(1)(B).

148. FED. R. CIV. P. 24(b)(3).

2. Standing

The doctrine of standing has proved extremely difficult for students, scholars, attorneys, and even the courts to understand and communicate.[149] Even the Supreme Court has inconsistently applied the doctrine throughout its history.[150] Possibly the clearest articulation of the standing doctrine occurred in 1976 when the Supreme Court decided *Simon v. Eastern Kentucky Welfare Rights Org.*[151] Growing out of Article III's case or controversy requirement,[152] the Court in *Simon* reiterated that standing "'focuses on the party seeking to get his complaint before a federal court and not on the issues he wishes to have adjudicated.'"[153] Specifically, standing asks "whether the plaintiff has alleged such a personal stake in the outcome of the controversy as to warrant his invocation of federal court jurisdiction and to justify exercise of the court's remedial powers on his behalf."[154] Ultimately, if parties lack standing before the court, then a proper case or controversy does not exist and the courts have "no business" hearing their dispute.[155]

In *Lujan v. Defenders of Wildlife*, the Supreme Court established a three-pronged test that a court should use to determine whether a party has sufficient standing to litigate before it.[156] First, the individual must have suffered an "injury in fact" that is "concrete and particularized" and "actual or imminent."[157] Second, the complained-of injury must be "fairly . . . trace[able] to the challenged action . . . and not . . . th[e] result [of] the independent action of some third party not before the court."[158] Third, the injury must be redressable by a favorable decision of that court.[159]

---

149. Valley Forge Christian Coll. v. Ams. United for Separation of Church & State, Inc., 454 U.S. 464, 475 (1982) ("We need not mince words when we say that the concept of 'Art. III standing' has not been defined with complete consistency in all of the various cases decided by this Court which have discussed it, nor when we say that this very fact is probably proof that the concept cannot be reduced to a one-sentence or one-paragraph definition.").

150. *See id.*

151. 426 U.S. 26, 37–38 (1976).

152. U.S. CONST. art. III, § 2, cl. 1.

153. *Simon*, 426 U.S. at 38 (quoting Flast v. Cohen, 392 U.S. 83, 99 (1968)).

154. *Simon*, 426 U.S. at 38 (quoting Warth v. Seldin, 422 U.S. 490, 498–99 (1975)) (internal quotation marks omitted).

155. DaimlerChrysler Corp. v. Cuno, 547 U.S. 332, 341 (2006).

156. *See* Lujan v. Defenders of Wildlife, 504 U.S. 555, 560 (1992) ("Over the years, our cases have established that the irreducible constitutional minimum of standing contains three elements.").

157. *Id.* (citation omitted).

158. *Id.* at 560–61 (alterations in original) (citation omitted).

159. *Id.* at 561 (citation omitted).

The doctrines of standing and intervention collided in 1986 when the Supreme Court decided *Diamond v. Charles*.[160] In *Diamond*, an Illinois doctor intervened in a district court proceeding in support of the defendant where the constitutionality of the state's abortion laws were being challenged.[161] The State of Illinois was not interested in pursuing an appeal, but Dr. Diamond was and subsequently filed a Notice of Appeal.[162] The Court held that because Diamond lacked his own Article III standing, it lacked jurisdiction for his appeal.[163] Specifically, Diamond's "status as an intervenor . . . whether permissive or as of right, does not confer standing sufficient to keep the case alive in the absence of the State on this appeal."[164] Importantly, however, the Court noted that had the State chosen to appeal, Diamond would have had the ability to "ride 'piggyback' on the State's undoubted standing."[165]

Shortly after *Diamond*, the Court established the *Bowsher* Doctrine to maintain that standing is met for *all* moving parties once *any one* moving party satisfies the three-pronged standard.[166] Then, in 2003, the Court specifically extended this principle to the concept of intervention as of right when it intentionally refused to address whether the intervening Members of Congress had their own Article III standing in *McConnell v. FEC*.[167] Of particular note in *McConnell*, though, is that the Members of Congress assumed an identical position as the original party with whom they sided.[168] One commentator has suggested that this fact, coupled with the *Diamond* holding, suggests that there are limits on an intervenor's freedom to act on its own.[169] What is ultimately important, however, is that an intervenor under Rule 24(a) who is advancing the same position as the original party with whom it has sided has no individual standing requirement.[170]

---

160. 476 U.S. 54 (1986).

161. *Id.* at 57–58.

162. *Id.* at 61.

163. *Id.* at 71.

164. *Id.* at 68.

165. *Id.* at 64.

166. *See* Bowsher v. Synar, 478 U.S. 714, 721 (1986) (emphasis added) ("A threshold issue is whether the Members of Congress, members of the National Treasury Employees Union, or the Union itself have standing . . . It is clear that members of the Union . . . [have] standing under § 274(a)(2) and Article III. We therefore need not consider the standing issue as to the Union or Members of Congress.").

167. McConnell v. FEC, 540 U.S. 93, 233 (2003) ("It is clear, however, that the Federal Election Commission (FEC) has standing, and therefore we need not address the standing of the intervenor-defendants, whose position here is identical to the FEC's.") (citations omitted).

168. *Id.*

169. Elizabeth Zwickert Timmermans, *Has the* Bowsher *Doctrine Solved the Debate?: The Relationship Between Standing and Intervention as of Right*, 84 NOTRE DAME L. REV. 1411, 1425 (2009).

170. *See, e.g.*, *McConnell*, 540 U.S. at 233; Diamond v. Charles, 476 U.S. 54, 63

Most of the Circuit Courts of Appeals have yet to mention or cite *McConnell*, which has yielded confusion and inconsistency at the circuit and district court levels.[171] Prior to *McConnell*, the Eighth and D.C. Circuits took the position that independent standing was required for intervention.[172] Interestingly, however, the D.C. Circuit has recognized that intervening as a plaintiff implicates the standing doctrine in different ways than intervening as a defendant.[173] Regardless of this distinction, the D.C. Circuit has required Article III standing in both circumstances.[174] Since *McConnell*, the Eighth Circuit has reasserted its position that standing is required by affirming a recent district court's order denying intervention due to lack of standing.[175] Similarly, the D.C. Circuit has cited its pre-*McConnell* precedent on the issue as it continues to require independent standing for intervenors.[176]

In contrast, the Second, Fourth, Fifth, Sixth, Ninth, and Eleventh Circuits did not require independent standing for intervention even before *McConnell*,[177] and the Tenth Circuit has subsequently agreed with them in

---

(1986); San Juan Cnty., Utah v. United States, 420 F.3d 1197, 1205 (10th Cir. 2005) ("[O]n many occasions the Supreme Court has noted that an intervenor may not have standing, but has not specifically resolved that issue, so long as another party to the litigation had sufficient standing to assert the claim at issue.").

171. *See* Timmermans, *supra* note 169, at 1428.

172. *See, e.g.*, Roeder v. Islamic Republic of Iran, 333 F.3d 228, 233 (D.C. Cir. 2003) ("[D]ecisions of this court hold an intervenor must also establish its standing under Article III of the Constitution."); Standard Heating & Air Conditioning Co. v. City of Minneapolis, 137 F.3d 567, 570 (8th Cir. 1998) ("Article III standing is a prerequisite for intervention.").

173. *Roeder*, 333 F.3d. at 233 ("As to intervention in the district court, requiring standing for an applicant wishing to come in on the side of a plaintiff who has standing runs into the doctrine that Article III is satisfied so long as one party has standing. Requiring standing of someone who seeks to intervene as a defendant runs into the doctrine that the standing inquiry is directed at those who invoke the court's jurisdiction.").

174. *See, e.g.*, Fund for Animals, Inc. v. Norton, 322 F.3d 728, 731–32 (D.C. Cir. 2003) (requiring Article III standing for intervening defendant); Cook v. Boorstin, 763 F.2d 1462, 1470 (D.C. Cir. 1985) (requiring an intervening plaintiff to have standing).

175. Med. Liab. Mut. Ins. Co. v. Alan Curtis LLC, 485 F.3d 1006, 1008–09 (8th Cir. 2007).

176. City of Naples Airport Auth. v. FAA, No. 03-1308, 2004 WL 1080160, at *1 (D.C. Cir. 2004) (citing *Roeder*, 333 F.3d at 233; *Norton*, 322 F.3d at 731–32).

177. *See, e.g.*, Grutter v. Bollinger, 188 F.3d 394, 398 (6th Cir. 1999) ("[A]n intervenor need not have the same standing necessary to initiate a lawsuit."); Ruiz v. Estelle, 161 F.3d 814, 832 (5th Cir. 1998) ("We find the better reasoning in those cases which hold that Article III does not require intervenors to possess standing."); Shaw v. Hunt, 154 F.3d 161, 165 (4th Cir. 1998) ("Key to our analysis is the Supreme Court's ruling that a party who lacks standing can nonetheless take part in a case as a permissive intervenor.") (citing SEC v. United States Realty & Improvement Co., 310 U.S. 434, 459 (1940)); Chiles v. Thornburgh, 865 F.2d 1197, 1213 (11th Cir. 1989) ("[A] party seeking to intervene need not demonstrate that he has standing in addition to meeting the requirements of Rule 24 as long as there exists a justiciable case and controversy between the parties already in the

light of *McConnell*.[178] These circuits have recognized that "Article III does not require each and every party in a case to have such standing."[179] Thus, the addition of a standingless intervenor would not destroy the court's ability to hear the case. Imperatively, these circuits distinguish between intervenors entering an active litigation and those where the potential intervenors would be "the only parties on appeal adverse to plaintiffs."[180] In the latter situation, both permissive and as-of-right intervenors must meet the requirements for Article III standing in addition to Rule 24.[181]

Finally, the First and Seventh Circuits have traditionally held that the interests required for intervention are equivalent to that of standing.[182]

---

lawsuit."); Portland Audubon Soc'y v. Hodel, 866 F.2d 302, 308, n.1 (9th Cir. 1989) (noting that, historically, it has "resolved intervention questions without making reference to standing doctrine"); United States Postal Serv. v. Brennan, 579 F.2d 188, 190 (2d Cir. 1978) ("The existence of a case or controversy having been established . . . there was no need to impose the standing requirement upon the proposed intervenor.").

Of note, the Fourth Circuit has chosen to differentiate standing requirements based on whether permissive or as-of-right intervention is being sought. *See Shaw*, 154 F.3d at 165. Citing precedent on the issue of what constitutes an adequate interest under Rule 24, it has interpreted the Supreme Court's statement that permissive intervention "plainly dispenses with any requirement that the intervenor shall have a direct personal or pecuniary interest in the subject of the litigation" to mean that Rule 24(b) does not require independent, Article III standing. *Id.*; *SEC*, 310 U.S. at 459.

178. San Juan Cnty. v. United States, 503 F.3d 1163, 1172 (10th Cir. 2007) (en banc), *vacating* 420 F.3d 1197 (10th Cir. 2005) ("On rehearing en banc we adopt the panel's reasoning on this issue and hold that parties seeking to intervene under Rule 24(a) or (b) need not establish Article III standing 'so long as another party with constitutional standing on the same side as the intervenor remains in the case.'") (quoting *San Juan Cnty.*, 420 F.3d at 1206).

179. *Ruiz*, 161 F.3d at 832.

180. Kootenai Tribe of Idaho v. Veneman, 313 F.3d 1094, 1109 (9th Cir. 2002); *see also* Associated Builders & Contractors v. Perry, 16 F.3d 688, 690 (6th Cir. 1994) ("An intervenor need not have the same standing necessary to initiate a lawsuit in order to intervene in an existing district court suit where the plaintiff has standing.").

181. *Veneman*, 313 F.3d at 1107 ("Before we address whether the district court erred in granting intervention under . . . [Rule] 24(b), we must first determine whether intervenors have Article III standing . . . [because] intervenors . . . [are] the only parties on appeal adverse to plaintiffs."); Didrickson v. United States Dep't of Interior, 982 F.2d 1332, 1337–38 (9th Cir. 1992) ("A permissive defendant-intervenor must have independent jurisdictional grounds on which to pursue an appeal, absent an appeal by the party on whose side the intervenor intervened.").

182. Cotter v. Mass. Ass'n of Minority Law Enforcement Officers, 219 F.3d 31, 34 (1st Cir. 2000) ("[I]n the ordinary case, an applicant who satisfies the 'interest' requirement of the intervention rule is almost always going to have a sufficient stake in the controversy to satisfy Article III as well.") (citing Transamerica Ins. Co. v. South, 125 F.3d 392, 396 n.4 (7th Cir. 1997)).

The First Circuit, however, did qualify this statement in *Cotter* by stating that "[s]tanding is an immensely complicated set of doctrines and it may be that there are unusual cases where an intervenor could satisfy the interest requirement of Rule 24(a)(2) without

Surprisingly, the Seventh Circuit has recently changed position on this issue and now distinguishes between intervenors looking to join a party already actively litigating an issue and those looking to continue litigation on their own.[183] Although standing is required in the latter scenario, "the predominant view is that intervention does not require that the intervenor have an interest sufficient under Article III" as long as the original parties are still litigating.[184] The Seventh Circuit has recognized several advantages that can result from allowing intervention without Article III standing, including a dispute "expedited or made more accurate or otherwise improved by allowing someone to enter the litigation, conduct discovery, examine and cross-examine witnesses, and otherwise disport himself as a party."[185]

Thus, patent attorneys attempting to intervene in district courts within the Fourth, Fifth, Sixth, Seventh, Ninth, Tenth, and Eleventh Circuits have a better chance of surviving opposing arguments demonstrating their lack of standing. On the other hand, the Eighth and D.C. Circuits, and potentially the First, will likely require the attempting intervenor to establish independent Article III standing. As a result, patent attorneys will have the burden to prove that a finding of inequitable conduct against them satisfies the three-pronged test for standing established in *Lujan*.[186] Many, including myself, believe that this is a nearly impossible burden to satisfy on the part of patent attorneys, particularly the injury-in-fact and traceability requirements.

3. Intervention Requirements: Permissive vs. As-of-Right

Rule 24(a)(2) requires an intervenor to show "an interest relating to the property or transaction which is the subject of the action and the applicant is so situated that the disposition of the action may as a practical matter impair or impede the applicant's ability to protect that interest."[187] In 1971, the Supreme Court interpreted this provision as requiring a "significantly protectable interest."[188] In *Donaldson v. United States*, the Court held that a taxpayer who was being investigated by the Internal Revenue Service could not intervene in an enforcement proceeding concerning Internal Revenue

---

having the stake in the controversy needed to satisfy Article III." *Id.* Because such circumstances were not present in *Cotter*, the First Circuit saw "no reason to concern ourselves with the abstract question whether an intervenor-defendant must show some separate form of standing." *Id.*

183. *See* Korczak v. Sedeman, 427 F.3d 419, 420–21 (7th Cir. 2005).

184. *Id.* at 421.

185. *Id.* at 422.

186. A discussion as to whether or not the standing requirements would be satisfied in such a situation is beyond the scope of this Comment.

187. FED. R. CIV. P. 24(a)(2).

188. Donaldson v. United States, 400 U.S. 517, 531 (1971).

summonses received by his employer "simply because it is his tax liability that is the subject of the summons."[189] Under this reasoning, one could infer that the burden may be too high for a patent attorney to intervene under this standard simply because the patent the attorney prosecuted is the subject of the infringement suit.[190]

Indeed, defining exactly what satisfies the "significantly protectable interest" standard has proven extremely difficult for the courts.[191] Although the Advisory Committee Notes on the 1966 Rule 24 Amendment counsel for a liberal, practical application of Rule 24(a),[192] not every court has acted accordingly. In fact, many circuits have endorsed a rigid standard requiring a Rule 24(a) interest to be "direct, substantial, and legally protectable"; but again, even this framework has been characterized as "problematic."[193] Frequently, district courts "pay lip service to the [direct, substantial, and legally protectable] test [and then] regularly manage to manipulate (ignore?) the language to reach the result required by practical considerations."[194] What is agreed upon, however, is that demonstrating the sufficiency of a Rule 24(a) "interest" requires a "highly fact-specific" inquiry.[195]

In addition to establishing the sufficiency of the intervenor's interest, the interest cannot already be "adequately represented by existing parties."[196] The Supreme Court addressed this issue in *Trbovich v. United Mine Workers*,[197] where a whistleblowing union member moved to intervene in the subsequent action brought by the Secretary of Labor under the Labor-Management Reporting and Disclosure Act of 1959 to invalidate the union's election of officers.[198] The Court held that because the Secretary's interests "transcend[ed] the narrower interest of the complaining union member" and included "an obligation to protect the 'vital public interest'" the union member may have a valid complaint about the performance of his lawyer, the Secretary of Labor.[199] Thus, "there is sufficient doubt about the adequacy of representation to warrant intervention."[200] "The requirement is satisfied if the applicant shows that

189. *Id.* at 530.

190. *See* Nisus Corp. v. Perma-Chink Sys., Inc., 497 F. 3d 1316, 1322 (Fed. Cir. 2007).

191. *See* 6 JAMES WM. MOORE ET AL., MOORE'S FEDERAL PRACTICE § 24.03[2][a], at 24–29 (3d ed. 2006) ("Courts have adopted a variety of approaches and a wide range of terminology in discussing the issue of interest.").

192. FED. R. CIV. P. 24(a), advisory committee's note (1966 Amendment).

193. San Juan Cnty.. v. United States, 503 F.3d 1163, 1192 (10th Cir. 2007).

194. *Id.* at 1193.

195. Utah Ass'n of Counties v. Clinton, 255 F.3d 1246, 1251 (10th Cir. 2001).

196. FED. R. CIV. P. 24(a).

197. Trbovich v. United Mine Workers, 404 U.S. 528 (1972).

198. *Id.* at 529–30.

199. *Id.* at 539 (quoting Wirtz v. Local 153, Glass Bottle Blowers Ass'n, 389 U.S. 463, 475 (1968)).

200. *Id.* at 538 (citation omitted).

the representation of his interest 'may be' inadequate; and the burden . . . should be treated as minimal."[201]

On the other hand, the Tenth Circuit has ruled that when an intervenor's *objectives* are shared by an existing party, "representation is adequate," regardless of either party's *motivations*.[202] These decisions seem to conflict with the Supreme Court's holding in *Trbovich* where the Secretary and the union member clearly shared the same objective—invalidating the election—yet the Secretary's representation may not have been adequate.[203] Clearly, patent attorneys seeking to intervene will share the same objective as the plaintiffs, to defeat the inequitable conduct charges, despite the fact that the parties' motivations for doing so may be entirely unrelated.

The Sixth Circuit has taken a stance on the issue of adequate representation which is particularly difficult for inventors.[204] In *Jordan v. Michigan Conference of Teamsters Welfare Fund*, participants in the Michigan Conference of Teamsters Welfare Fund ("MCTWF") sued the MCTWF in connection with the administration and distribution of healthcare and welfare benefits.[205] During the proceeding, the court ruled that any attorneys' fees received by the plaintiffs in the proceeding that were transferred to the International Brotherhood of Teamsters AFL-CIO ("IBT") would constitute a prohibited transaction under ERISA.[206] IBT sought to intervene under Rule 24(a) in order to pursue an appeal to recover the money it had advanced to plaintiffs' counsel.[207] The Sixth Circuit assessed whether IBT's interests were adequately represented and concluded that IBT's only argument, that it "would be more vigorous in pursuing its claim for reimbursement than Plaintiffs . . . . [, failed] to identify any potential inadequacy in Plaintiff's continued representation of the IBT's interests."[208]

---

201. *Id.* at 538 n.10.

202. City of Stilwell, Okla. v. Ozarks Rural Elec. Co-op., 79 F.3d 1038, 1042 (10th Cir. 1996).

203. *Trbovich*, 404 U.S. at 529–30.

204. *See* Jordan v. Mich. Conf. of Teamsters Welfare Fund, 207 F.3d 854, 863 (6th Cir. 2000).

205. *Id.* at 857.

206. *Id.*

207. *Id.* at 858.

208. *Id.* at 863. The court stated further:

> [T]his Court has held that a movant fails to meet his burden of demonstrating inadequate representation when 1) no collusion is shown between the existing party and the opposition; 2) the existing party does not have any interests adverse to the intervener [sic]; and 3) the existing party has not failed in the fulfillment of its duty.

Id. (citing Bradley v. Miliken, 828 F.2d 1186, 1192 (6th Cir. 1987)).

Permissive intervention is governed by Rule 24(b) and requires only that the intervenor have "a claim or defense that shares with the main action a common question of law or fact."[209] Notions of sufficient interest and inadequate representation are ignored pursuant to Supreme Court precedent that states "permissive intervention plainly dispenses with any requirement that the intervenor shall have a direct personal or pecuniary interest in the subject of the litigation."[210] Upon a timely application for permissive intervention by an individual whose claim or defense shares a common question of law or fact, "the requirement of the rule has been satisfied and it is then discretionary with the court whether to allow intervention."[211]

Frequently, factors of equity can influence a determination of whether permissive intervention is appropriate. The Ninth Circuit in *Kootenai v. Veneman* allowed permissive intervention in light of the case's magnitude and the potential intervenor's contribution "to the equitable resolution of [the] case."[212] Characterizing the reasoning as "good and substantial," the Ninth Circuit noted that "[u]nder these circumstances it is clear . . . that the presence of intervenors would assist the court in its orderly procedures leading to the resolution of this case, which impacted large and varied interests. The district court did not abuse its discretion in granting the intervenors permissive intervention under Rule 24(b)."[213]

Finally, undue delay or prejudice to the original parties must be considered in a permissive intervention inquiry.[214] Frequently, as long as the motion for intervention is filed near the outset of the proceedings and the intervenor requires no additional time for briefing and scheduling, the court will rarely fine undue delay and prejudice.[215]

Patent attorneys attempting to intervene in a district court in light of inequitable conduct allegations will always have a "defense that shares with the main action a common question of law or fact."[216] Patent owners will have to defend against the same charges in order to enforce their rights. Therefore, a timely application for permissive intervention would seem to satisfy all requirements and would permit the patent attorney to defend herself.

On the other hand, Rule 24(a) presents several substantial hurdles that will frustrate a patent attorney's efforts to intervene as-of-right. Aside from any standing requirements exclusive for as-of-right intervention,[217] the

---

209. FED. R. CIV. P. 24(b)(2).

210. SEC v. United States Realty & Improvement Co., 310 U.S. 434, 459 (1940).

211. Kootenai Tribe of Idaho v. Veneman, 313 F.3d 1094, 1111 (9th Cir. 2002).

212. *Id.*

213. *Id.*

214. FED. R. CIV. P. 24(b)(2).

215. *See Veneman*, 313 F.3d at 1111 n.10 (9th Cir. 2002).

216. FED. R. CIV. P. 24(b)(2).

217. *See* Shaw v. Hunt, 154 F.3d 161, 165 (4th Cir. 1998) ("Key to our analysis is the Supreme Court's ruling that a party who lacks standing can nonetheless take part in a case as

patent attorney must convince the court that their professional reputation qualifies as a direct and substantial legal interest.[218] Furthermore, the attorney must persuade the judge that the patent owner is not already adequately protecting that interest.[219] In the Sixth and Tenth Circuits, this will be extremely difficult to do because both the patent attorney and the patent owner share the same objective: to defeat the inequitable conduct allegations.[220] However, the Supreme Court's insistence that the rule only requires that the representation of the interest "may be" inadequate[221] could be satisfied if a finding of inequitable conduct includes multiple named individuals. In such a case, the patent attorney's particular interest will not be to defeat the charges altogether but rather to specifically clear the attorney's name.

4. Timeliness

Timeliness is required for both permissive and as-of-right intervention.[222] Untimely applications for intervention *must* be denied.[223] Although this requirement is mandatory, and therefore typically treated as a threshold issue, "timeliness is to be determined from all circumstances . . . by the [district] court in the sole exercise of its sound discretion."[224] Factors that have been developed to assess timeliness include the point to which the suit has progressed,[225] the applicant's purpose for seeking intervention,[226] the amount of time the applicant knew or reasonably should have known of the applicant's interest in the case,[227] the prejudice to the original parties,[228] and any mitigating circumstances.[229] The courts pay particular attention to when the applicant first learned of its interest in the case, and how much time elapsed before the applicant sought to intervene.[230] Generally, an

---

a permissive intervenor.").

218. San Juan Cnty. v. United States, 503 F.3d 1163, 1192 (10th Cir. 2007).

219. FED. R. CIV. P. 24(a)(2).

220. *See* Jordan v. Mich. Conf. of Teamsters Welfare Fund, 207 F.3d 854, 863 (6th Cir. 2000); City of Stilwell v. Ozarks Rural Elec. Co-op. 79 F.3d 1038, 1042 (10th Cir. 1996).

221. Trbovich v. United Mine Workers, 404 U.S. 528, 538 n.10 (1972).

222. FED. R. CIV. P. 24(a), (b); *see also* NAACP v. New York, 413 U.S. 345, 365 (1973) ("Whether intervention be claimed of right or as permissive, it is at once apparent, from the initial words of both Rule 24(a) and Rule 24(b), that the application must be 'timely.'").

223. *See NAACP*, 413 U.S. at 365.

224. *Id.* at 366.

225. *Id.* at 365–66.

226. Hodgson v. United Mine Workers, 473 F.2d 118, 129 (D.C. Cir. 1972).

227. Stallworth v. Monsanto Co., 558 F.2d 257, 264 (5th Cir. 1977).

228. Culbreath v. Dukakis, 630 F.2d 15, 21 (1st Cir. 1980).

229. *Stallworth*, 558 F.2d at 266; *see also*, Jordan v. Mich. Conf. of Teamsters Welfare Fund, 207 F.3d 854, 862 (6th Cir. 2000) (applying all the timeliness factors).

230. *See Jordan*, 207 F. 3d at 862. The Sixth Circuit noted that the applicant knew of its

applicant "must move to protect its interest no later than when it gains some actual knowledge that a measurable risk exists."[231]

If the district court has already entered a final judgment and dispensed with the action, intervening in the district court is not possible.[232] Even if the attorney responsible for the inequitable conduct finding filed a Rule 60(b) motion for reinstatement of the judgment, the applicant would still need to apply for intervention in the district court to pursue the motion.[233] When this occurred in the Fifth Circuit, the court held that intervention was improper because the suit had been dismissed, and as a result, there was no existing case or controversy within the court's jurisdiction.[234]

All patent appeals go to the Federal Circuit. Having missed the opportunity to intervene in the district court, a patent attorney who has just learned of the inequitable conduct findings against him or her may attempt to intervene in the appeal. Federal Rule Appellate Procedure 15(d) provides for such intervention in an appeal.[235] Furthermore, the Supreme Court has stated that "the policies underlying intervention may be applicable in appellate courts."[236] After considering these policies, the Court stated in *International Union v. Scofield* that the applicant could intervene permissively or as-of-right in the Seventh Circuit review of a Labor Board decision.[237]

---

interest in the suit prior to when the defendants (who opposed the motion) learned of it and that applicant had a six-month window from when it knew of its interests until the district court rendered its decision affecting that interest. *Id.* IBT moved to intervene following the district court's decision. *Id. See also* Banco Popular de Puerto Rico v. Greenblatt, 964 F.2d 1227, 1231 (1st Cir. 1992) ("[C]ourts have historically viewed post-judgment intervention with a jaundiced eye in situations where the applicant had a reasonable basis for knowing, before judgment, that its interest was at risk.").

231. *Banco Popular*, 964 F.2d at 1231 (citing *Culbreath*, 630 F.2d at 21, and Alaniz v. Tillie Lewis Foods, 572 F.2d 657, 659 (9th Cir. 1978) (per curiam), *cert. denied*, 439 U.S. 837 (1978)).

232. Smith v. Mo. Pac. R.R., 615 F.2d 683, 685 (5th Cir. 1980).

233. *See* Ericsson, Inc. v. Interdigital Comm. Corp., 418 F.3d 1217, 1224 (5th Cir. 2005) ("The plain language of Rule 60(b) only allows relief to be given to a 'party' to the litigation.").

234. *Id.* at 1221–22 (citing Kendrick v. Kendrick, 16 F.2d 744, 745 (5th Cir. 1927)).

235. FED. R. APP. PRO. 15(d). The rule states:

> Unless a statute provides another method, a person who wants to intervene in a proceeding under this rule must file a motion for leave to intervene with the circuit clerk and serve a copy on all parties. The motion or other notice of intervention authorized by statute must be filed within 30 days after the petition for review is filed and must contain a concise statement of the interest of the moving party and the grounds for intervention.

Id.

236. Int'l Union v. Scofield (*Auto Workers*), 382 U.S. 205, 217 n.10 (1965).

237. *Id.*

Importantly, the Supreme Court has held that Federal Rule of Appellate Procedure 3(c), which requires that the Notice of Appeal "specify the party or parties taking the appeal by naming each one,"[238] is a jurisdictional requirement.[239] As a result, the Second Circuit has denied an appeal by an individual who was not named as an appellant.[240] However, attorneys who have been chastised or sanctioned during the district court proceedings *and specified as an appellant* in the Notice of Appeal, have been allowed to intervene in appeals in the Second Circuit.[241]

Both the Federal Circuit and its predecessor court, the Court of Customs and Patent Appeals ("CCPA"), have allowed parties absent in the district court proceedings to intervene on appeal.[242] For example, in the case of *In re Echostar Communications Corp.*, the Federal Circuit allowed a non-participating law firm to intervene in the defendant's petition for a writ of mandamus.[243] Although the Federal Circuit neither commented on the applicable procedural law of the Fifth Circuit, where the case originated, nor discussed why leave for intervention was granted, the court did consider the law firm's arguments on the issue.[244] The CCPA permitted the assignee of a patent, who was not a party in the district court, to intervene in the appeal over a Trade Commission order that invalidated the patent.[245] Further, the court still allowed intervention, even in light of the fact that the applicant's motion was three days late.[246]

In *Elliot Industries v. BP American Production Co.*, the Tenth Circuit allowed a party to intervene as of right on appeal in support of the appellees to challenge subject matter jurisdiction.[247] Recognizing that the factors involved in Rule 24 are applicable to intervention on appeal, the court allowed the applicant to intervene because none of the other parties were challenging subject matter jurisdiction.[248] The Tenth Circuit reasoned that its inquiry into the essential element of subject matter jurisdiction would be "aided by the presence of an interested party like [the applicant]" and that

238. FED. R. APP. PRO. 3(c).

239. Torres v. Oakland Scavenger Co., 487 U.S. 312, 314 (1988).

240. DeLuca v. Long Island Lighting Co., 862 F.2d 427, 429–30 (2d Cir. 1988).

241. Fromson v. Citiplate, Inc., 886 F.2d 1300, 1304 (Fed. Cir. 1989); Penthouse Int'l, Ltd. v. Playboy Enters., 663 F.2d 371, 373 (2d Cir. 1981).

242. *See In re* EchoStar Comm. Corp., 448 F.3d 1294, 1296–97 (Fed. Cir. 2006); Canadian Tarpoly Co. v. United States Int'l Trade Comm'n, 649 F.2d 855, 857 (C.C.P.A. 1981).

243. *EchoStar Comm. Corp.*, 448 F.3d at 1296–97.

244. *See id.* at 1296–1305.

245. *Canadian Tarpoly Co.*, 649 F.2d at 856.

246. *Id.*

247. Elliot Indus. v. BP Am. Prod. Co., 407 F.3d 1091, 1103–04 (10th Cir. 2005).

248. *Id.* at 1103 ("[Applicant's] interest may be harmed if [he] is not permitted to intervene on appeal.").

any additional delay or prejudice would be "minimal compared with the importance of addressing the question."[249]

There is a "well established general rule . . . that a court of appeals will not consider an issue raised for the first time on appeal."[250] This rule has strong implications for patent attorneys attempting to intervene in an appeal, as in *Therasense*.[251] With the scope of review limited to what was before the district court, a patent attorney intervening in an appeal will be constrained to argue only issues that were presented to the district court. Thus, if only issues argued by the original parties in the district court can be reviewed, one might ask what the intervention by the patent attorney actually contributes to the appeal.

Without the ability to introduce new evidence or issues in front of the appellate court, what does the intervening attorney contribute to litigation? Ultimately, the inclusion of the patent attorney in the litigation, even if the attorney is limited in only arguing issues before the district court, is beneficial to the fair and equitable distribution of justice. The perspectives advanced by the patent attorney, while conforming with the scope of appellate review, can still differ from those raised by the original parties. The question should not be what does the intervention contribute to the litigation, but whether intervention is allowed under the rules and whether it contributes to the "equitable resolution of [the] case."[252]

5. *Nisus Corp.* and *Therasense*

Michael Teschner moved to intervene in the district court proceedings in *Nisus Corp.*[253] His motion was denied because it was untimely, having been filed after the entry of judgment in the case.[254] Additionally, the original parties were not pursuing an appeal.[255] The Federal Circuit affirmed the denial of the motion to intervene stating, "Mr. Teschner lacks a substantial legal interest in the underlying litigation."[256] The Federal Circuit dismissed Attorney Teschner's argument that *Penthouse International v. Playboy Enterprises* supported his position because the attorney in

249. *Id.* at 1103–04.

250. Cherry v. City of New York, No. 09-2581-CV, 2010 WL 2501040, at *2 (2d Cir. June 18, 2010) (citing Singleton v. Wulff, 428 U.S. 106, 120–21 (1976)). An exception to this general rule allows courts to permit the introduction of a new issue on appeal to "avoid obvious or manifest injustice." *Id.* (citing Thomas E. Hoar, Inc. v. Sara Lee Corp., 900 F.2d 522, 527 (2d Cir.1990)).

251. Pope's Motion to Intervene, *supra* note 1, at 1.

252. Kootenai Tribe of Idaho v. Veneman, 313 F.3d 1094, 1111 (9th Cir. 2002).

253. Nisus Corp. v. Perma-Chink Sys., Inc., No. 3:03-CV-120, 2006 WL 2128903, at *1 (E.D. Tenn. July 27, 2006).

254. *Id.* at *10.

255. *See id.*

256. Nisus Corp. v. Perma-Chink Sys., Inc., 497 F.3d 1316, 1322 (Fed. Cir. 2007) (citation omitted).

Penthouse International had been sanctioned by the court and ordered to pay a fine.[257] Attorney Teschner was facing no such fine or sanction.[258] Further, the court opined that "[t]here is no reason to believe Sixth Circuit law would permit Mr. Teschner to intervene for purposes of pursing an appeal in the present case under these circumstances."[259]

Lawrence Pope also moved to intervene in *Therasense* following the district court's entry of a final judgment.[260] He filed his motion with the Federal Circuit after the original parties to the lawsuit filed their notice of appeal.[261] Unlike the circumstances in *Nisus Corp.*, where the parties were not pursuing an appeal, the parties in *Therasense* certainly were.[262] However, relying almost entirely on *Nisus Corp.*, the Federal Circuit denied in a one-page order Attorney Pope's motion to intervene, stating that because Pope had not been formally sanctioned, he lacked a substantial interest in the appeal.[263] The court ignored the fact that the parties were already pursuing an appeal in the case, even after Pope expressly made that argument in his brief.[264]

In Attorney Pope's motion, he cited *Penthouse International* to support the fact that there was precedent for allowing attorneys to intervene in an appeal.[265] However, the Federal Circuit stated that his "reliance on *Penthouse Int'l* [was] misplaced."[266] The court "expressly distinguished *Penthouse Int'l* because it involved an attorney who sought to intervene after a district court sanctioned the attorney and directed the attorney to pay a sanction."[267] The court did not cite to any Ninth Circuit case law in support of this distinction.[268] In fact, by citing *Nisus Corp.*, the Federal Circuit actually relied on its interpretation of Sixth Circuit law to decide Pope's motion.[269] This is puzzling in that the Federal Circuit decides

---

257. *Id.*

258. *Id.*

259. *Id.* at 1323 n.1 (citation omitted).

260. *See* Pope's Motion to Intervene, *supra* note 1, at 1.

261. *See id.*

262. Pope's Reply in Support of His Motion for Leave to Intervene on Appeal at 5–6, Therasense, Inc. v. Becton, Dickinson & Co., 593 F.3d 1289 (Fed. Cir. 2008) (Nos. 2008-1511, -1512, -1513, -1514, -1595), [hereinafter Pope's Reply] *available at* http://www.patentlyo.com/lawrence_20s._20pope_s_20reply_20in_20support_20of_20his_20motion_20for_20leave_20to_20intervene_20on_20appeal.pdf (citations omitted); *see also* Nisus Corp. v. Perma-Chink Sys., Inc., 497 F.3d 1316, 1318 (Fed. Cir. 2007).

263. *See Therasense* Order, *supra* note 10.

264. *See id.* at 2; Pope's Motion to Intervene, *supra* note 1, at ¶ 18, 20.

265. See Pope's Motion to Intervene, *supra* note 1, at ¶ 23 (citations omitted).

266. *Therasense* Order, *supra* note 10, at 2; see Pope's Motion to Intervene, *supra* note 1, at ¶ 23 (citations omitted).

267. *Therasense* Order, *supra* note 10, at 2.

268. *See id.*

269. *See* Nisus Corp. v. Perma-Chink Sys., Inc. 497 F.3d 1316, 1323 n.1 (Fed. Cir. 2006).

matters "not unique to patent law" according to the "law of the appropriate regional circuit."[270] Intervention issues are hardly unique to patent law.[271]

Because so many different issues influence whether intervention is warranted, and that the circuits differ in their approach to each issue, the law surrounding the ability of patent attorneys to intervene where they are the subject of inequitable conduct claims is entirely unpredictable.[272] Unpredictability is unfortunate for a regime that is largely dependent upon the application of a universal standard.

B. Appeal the District Court Opinion

Consider the situation where the patent attorney is not notified of the district court proceedings, and therefore, does not have the opportunity to intervene. If the patent owner does not elect to call the attorney as a witness, the district court's opinion is brought to the patent attorney's attention only after the slip opinion is released. Furthermore, the original parties may instead reach a settlement that avoids appellate litigation on the issue. What are the options for the aggrieved patent attorney at this point? Instead of just waiting around for the PTO to call, the attorney may choose to file his own appeal in the Federal Circuit.

The standing doctrine will be a significant obstacle that the patent attorney must also overcome to appeal the district court's opinion.[273] As the Supreme Court stated in *Diamond v. Charles*, those who lack Article III standing, fail to satisfy the jurisdictional requirement for appeal.[274] Another major hurdle could be that the patent attorney may not appeal a judgment, decree, or order of the district court.[275] With few exceptions,[276] Congress granted the courts of appeals jurisdiction only over "final decisions" by the district courts.[277]

---

270. Ericsson, Inc. v. Interdigital Commc'n Corp., 418 F.3d 1217, 1220–21 (Fed. Cir. 2005).

271. Haworth, Inc. v. Steelcase, Inc., 12 F.3d 1090, 1092 (Fed. Cir. 1993).

272. *See infra* Part IV.

273. *See supra* Part IV(A)(3).

274. *See* Diamond v. Charles, 476 U.S. 54, 71 (1986).

275. For example, the district court could have found that several individuals committed inequitable conduct and that they each contributed to the unenforceability of the patent. In that scenario, a particular patent attorney may only care to appeal the decision as it pertains to him, not to the patent as a whole. Therefore, only part of the opinion would be at issue on appeal, not the district court's final decision. *See* Black v. Cutter Labs, 351 U.S. 292, 297 (1956) ("This Court, however, reviews judgments, not statements in opinions.").

276. *See* 28 U.S.C. § 1292 (2006).

277. *Id.* § 1291; *see also id.* § 2106 (Appellate courts "may affirm, modify, vacate, set aside or reverse any judgment, decree, or order of a court lawfully brought before it for review.").

The closest applicable case law to an attorney appealing an inequitable conduct finding deals with appealing sanctions.[278] After introducing the "nonparty rule" and its limited exceptions, a brief discussion will ensue concerning the different types of sanctions handed down by a court. Next, this Comment will identify the different approaches taken by the circuit courts on the issue. Finally, it will explain how this issue was implicated in both *Nisus Corp.* and *Therasense.*

1. The "Nonparty Rule" and Types of Sanctions

The "nonparty rule" prevents those who were not a party to the litigation from appealing its judgment.[279] Courts have developed this rule using both the standing doctrine and the finality requirement of 28 U.S.C. § 1291.[280] For example, the Tenth Circuit denied an attorney's appeal of a district court sanction against him because it "was not sufficiently injurious to confer standing."[281] Likewise, the First Circuit has also denied an appeal of a district court opinion because a final decision had not been rendered regarding the appealing attorney.[282] Of course, as with any other legal doctrine, limited exceptions have been carved out of this approach.[283]

Sanctions have been classified as traditional or informal.[284] Traditional sanctions include holding an attorney in contempt, assessing fines, or disqualification.[285] Often rule based, traditional sanctions "are designed primarily to ensure the integrity of the system and regulate conduct before

278. *See generally* David A. Simon, *Mo' Money, Mo' Problems: Should Appellate Courts Have Nonparty Jurisdiction Over Lawyers' Appeals From Nonmonetary Sanctions?*, 78 U. CIN. L. REV. 183 (2009) (discussing appellate court jurisdiction over nonparty lawyers' appeals from nonmonetary sanctions).

279. *See* Karcher v. May, 484 U.S. 72, 77 (1987) ("[W]e have consistently applied the general rule that one who is not a party or has not been treated as a party to a judgment has no right to appeal therfrome [sic].").

280. *See* 28 U.S.C. § 1291 (2006); *see, e.g.*, Teaford v. Ford Motor Co., 338 F.3d 1179 (10th Cir. 2003).

281. *Teaford*, 338 F.3d at 1181; *see also* Bolte v. Home Ins. Co., 744 F.2d 572 (7th Cir. 1984) (denying appeal based on 28 U.S.C. § 1291).

282. *See In re* Williams, 156 F.3d 86, 92 (1st Cir. 1998).

283. *See* 15A CHARLES ALAN WRIGHT, ARTHUR R. MILLER & EDWARD H. COOPER, FEDERAL PRACTICE AND PROCEDURE § 3902.1 (2d ed. 2010) ("Appeals have been permitted on showings that range from easy cases in which a nonparty is formally addressed by court order through less clear cases in which a nonparty is significantly affected."); *see also* Nisus Corp. v. Perma-Chink Sys, Inc., 497 F.3d 1316, 1319 (Fed. Cir. 2007) ("As an exception to tha general rule, a nonparty such as an attorney who is held in contempt or otherwise sanctioned by the court in the course of litigation may appeal from the order imposing sanctions.").

284. Judith A. McMorrow et al., *Judicial Attitudes Toward Confronting Attorney Misconduct: A View from the Reported Decisions*, 32 HOFSTRA L. REV. 1425, 1452 (2004).

285. *Id.* at 1452–53.

the court."[286] On the other hand, informal sanctions do not rely on statutes or rules and "include a court's decision to issue an opinion, naming the recalcitrant attorney, outlining his or her misdeeds in detail, and describing the court's disappointment and outrage."[287] Furthermore, scholars McMorrow, Gardinia, and Ricciardone argue that informal sanctions are more efficient than their traditional counterparts because they do not adhere to due process requirements.[288]

Arguably, informal sanctions are much more serious and damaging to a lawyer's professional career than traditional sanctions.[289] Unlike mere fines, informal sanctions, in the way of judicial criticism in the opinion, announce the judge's anger and outrage to the public.[290] As one commentator puts it, "[j]udges are prone to forget the sting of public criticism delivered from the bench. Such criticism, while potentially constructive, can also damage a lawyer's reputation and career."[291]

2. Circuit Confusion: Appealing Sanctions

The courts of appeals have not taken a consistent stance on whether attorneys may appeal district court sanctions against them, either traditional or informal.[292] Indeed, one scholar has characterized the circuits as "confused, rather than split, over the correct legal standard" for what types of sanctions are appealable.[293] Circuits have predominantly taken one of three positions: (1) that only monetary sanctions are appealable; (2) that sanctions must amount to explicit, formal reprimand to be appealable; or (3) that all sanctions are appealable as long as they are likely to damage an attorney's professional reputation.[294]

---

286. *Id.* at 1453.

287. *Id.*

288. *Id.*

289. *See* William W. Schwarzer, *Sanctions Under the New Federal Rule 11—A Closer Look*, 104 F.R.D. 181, 202 (1985).

290. *See* McMorrow, *supra* note 284, at 1454.

291. Schwarzer, *supra* note 289 at 201; *see also* Precision Specialty Metals, Inc. v. United States, 315 F.3d 1346, 1353 (Fed. Cir. 2003) ("A lawyer's reputation is one of his most important professional assets. Indeed, such a reprimand may have a more serious adverse impact upon a lawyer than the imposition of a monetary sanction.").

292. *See* Simon, *supra* note 278, at 186.

293. *Id.*

294. *See* Carla R. Pasquale, *Scolded: Can an Attorney Appeal a District Court's Order Finding Professional Misconduct?*, 77 FORDHAM L. REV. 219, 228–29 (2008); *cf.* Butler v. Biocore Med. Tech., Inc., 348 F.3d 1163, 1167 (10th Cir. 2003) (categorizing the positions as "never appealable, always appealable, and appealable only if the order is 'expressly identified as a reprimand'"); Simon, *supra* note 278, at 193 (categorizing the approaches as "(a) the all-or-nothing approach, (b) the formalist approach, and (c) the functionalist approach").

The harshest circuit on the issue of appealing sanctions is the Seventh, which mandates that only monetary sanctions from its district courts may be appealed.[295] In *Bolte v. Home Ins. Co.*, the judge for the U.S. District Court for the Western District of Wisconsin found that an attorney's conduct of concealing inconsistent witness statements from the other party was "reprehensible" and included this finding in the opinion.[296] The Seventh Circuit denied the appeal on the ground that 28 U.S.C. § 1291 was not "meant to allow people who were not even parties to a lawsuit in the district court to appeal from a wounding, critical, or even palpably injurious comment or finding by a district judge."[297] Likewise, in *Seymour v. Hug*, the court explained that only district court orders imposing monetary sanctions on the attorney could be appealed.[298]

The First, Ninth, and Federal Circuits allow attorneys to appeal formal, explicit reprimands only.[299] Under this approach, monetary sanctions as well as official reprimands may be appealed, but scathing commentary in the opinion may not be appealed.[300] For instance, in *In re Williams*, the First Circuit stated that "a jurist's derogatory comments about a lawyer's conduct, without more, do not constitute a sanction."[301] However, the court qualified this statement by insisting that "[w]ords alone may suffice if they are expressly identified as a reprimand."[302] One year after *Williams*, the Ninth Circuit followed suit and adopted the same approach.[303]

However, the Ninth Circuit has subsequently softened its stance on the issue.[304] In *United States v. Talao*, it maintained that "routine judicial commentary" was not appealable, but prescribed that a finding that an attorney willfully violated an ethical rule "per se, constitutes a sanction," accompanied by the requisite degree of formality to appeal.[305] Ultimately, the Ninth Circuit may now stand alone in a fourth approach that sits somewhere in between that of only allowing appeals from formal, explicit reprimands and that of allowing any professionally damaging commentary to be appealed.[306]

---

295. *See* Seymour v. Hug, 485 F.3d 926 (7th Cir. 2007); Bolte v. Home Ins. Co., 744 F.2d 572 (7th Cir. 1984).

296. *Bolte*, 744 F.2d at 572.

297. *Id.* at 573.

298. *Seymour*, 485 F.3d at 929.

299. *See* Pasquale *supra* note 294, at 228–29.

300. *See In re* Williams, 156 F.3d 86, 90 (1st Cir. 1998).

301. *Id.* at 92.

302. *Id.*

303. *See* Weissman v. Quail Lodge, Inc., 179 F.3d 1194, 1200 (9th Cir. 1999) ("We agree with the holding of Williams that words alone will constitute a sanction only 'if they are expressly identified as a reprimand.'") (quoting *Williams*, 156 F.3d at 93).

304. *See* United States v. Talao, 222 F.3d 1133 (9th Cir. 2000).

305. *Id.* at 1138.

306. If *Nisus Corp.* had occurred in the Ninth Circuit like *Therasense*, then would the court have felt that committing inequitable conduct would amount to an ethical rule violation

The Federal Circuit has also taken the stance that only formal reprimands may be appealed.[307] In *Precision Specialty Metals, Inc. v. United States*, the court held that a finding of a Rule 11 violation, without a monetary sanction, was appealable due to the formality of the Rule.[308] Interestingly, however, the court also specifically stated that a formal reprimand, disclosed in an unpublished opinion, "may have a more serious adverse impact upon a lawyer than the imposition of a monetary sanction."[309] Nonetheless, in *Nisus Corp.*, the Federal Circuit has since clarified that only "'sanctions or findings' [pursuant] to the formal imposition of the court's inherent power to penalize those who appear before it" were appealable.[310]

The Second, Fifth, Tenth and D.C. Circuits share the approach that all orders finding attorneys guilty of professional misconduct—and thus likely to damage their professional reputations—are appealable.[311] "[This] approach rejects the strict standard that looks to the form of the sanction and instead examines the substance of the alleged sanction itself. This approach asks whether the judicial decision acts as a formal sanction, rather than questioning whether the judge identified or intended it as one."[312] Furthermore, the position taken by these circuits also recognizes "the importance of an attorney's professional reputation"[313] and the practical effects of scathing commentary in a district court opinion.[314]

In *Sullivan v. Committee on Admission and Grievances*, the D.C. Circuit considered the appealability of a district court opinion stating that an attorney in the case had violated particular Canons of Ethics.[315] The D.C. Circuit concluded that because the district court's determination "reflect[ed] adversely on his professional reputation," the attorney could appeal the district court's determination that the attorney had violated certain Canons

---

and thus a "per se . . . sanction"? *Id.* Because Model Rule of Professional Conduct 3.3(a) requires candor toward a tribunal, there is certainly reason to believe that it would have. M.R.P.C. 3.3(a) (2010). Therefore, Attorney Teschner likely would have been able to have his day in court.

307. *See* Precision Specialty Metals, Inc. v. United States, 315 F.3d 1346 (Fed. Cir. 2003).

308. *Id.* at 1352.

309. *Id.* at 1353; *see also*, Pasquale, *supra* note 294, at 242 (discussing the irony of "the Federal Circuit['s] center[ing] its decision around the formality of the . . . reprimand, [while] it also heavily emphasized the negative effects that a court's admonishment can have on an attorney's professional reputation.").

310. Nisus Corp. v. Perma-Chink Sys., Inc., 497 F.3d 1316, 1321 (Fed. Cir. 2007) (citation omitted).

311. *See* Simon, *supra* note 278, at 197.

312. *See id.* (emphasis omitted).

313. Walker v. City of Mesquite, 129 F.3d 831, 832–33 (5th Cir. 1997).

314. *See id.*

315. 395 F.2d 954, 956 (D.C. Cir. 1967).

of Ethics.[316] Similarly, the Second Circuit reasoned that a trial court's finding of an ethics violation and its subsequent referral of the matter to the state disciplinary committee was "in the nature of a sanction," and allowed the appeal.[317]

Likewise, in *Walker v. City of Mesquite*, the Fifth Circuit held that a trial court's finding of an attorney's "blatant misconduct" at trial was appealable.[318] The Circuit was particularly concerned with the effects on the attorney's career, mentioning that the trial court's finding would "be seen as a blot on [the attorney's] professional record with a potential to limit his advancement in governmental service and impair his entering into otherwise inviting private practice."[319] Curiously, in the same breath the court stated that "the importance of an attorney's professional reputation, and the imperative to defend it when necessary, obviates the need for a finding of monetary liability or other punishment as a requisite for the appeal of a court order finding professional misconduct."[320] Therefore, the Fifth Circuit purports to require a formal reprimand in the way of either a monetary sanction or "other punishment" in order to appeal a judicial sanction, but then sets practical effects to constitute "other punishment."[321]

The Tenth Circuit sided with the majority of the other jurisdictions in *Butler v. Biocore Medical Technologies, Inc.*[322] After the district court found that Butler had committed ethics violations, a copy of the order was mailed to every courthouse in which Butler was admitted to practice.[323] The court recognized that "an order thus affecting an attorney's professional reputation imposes a legally sufficient injury to support appellate jurisdiction," and allowed Butler's appeal for several reasons.[324] Among them was a concern about the lack of an adversarial appeal that would result without including the attorney and a belief that excessive appellate litigation would be harnessed by the deferential standard of review applied.[325] Furthermore, the Tenth Circuit stated that "a rule requiring an explicit label as a reprimand ignores the reality that a finding of misconduct damages an attorney's reputation regardless of whether it is labeled as a reprimand and, instead, trumpets form over substance."[326]

---

316. *Id.*
317. *In re* Goldstein, 430 F.3d 106, 111 (2d Cir. 2005).
318. 129 F.3d 831, 832–33 (5th Cir. 1997).
319. *Id.* at 832.
320. *Id.*
321. *See id.* at 832–33.
322. 348 F.3d 1163, 1168 (10th Cir. 2003).
323. *Id.* at 1167.
324. *Id.*
325. *Id.* at 1168–69.
326. *Id.* at 1169.

3. *Nisus Corp.*

In *Nisus Corp.*, Michael Teschner filed a notice of appeal after the district court denied both his motion to intervene and his motion to amend the judgment.[327] The Federal Circuit began its analysis by deciding whether it had jurisdiction for the appeal pursuant to 28 U.S.C. § 1295.[328] Because of the jurisdictional nature of the analysis under § 1295, the court chose to "resolve the questions" by applying Federal Circuit law, and not the law of the regional circuit "from which the case arose."[329] After pointing out the general exclusion of non-party appeals, the court explored the varying approaches on the scope of the exceptions to that rule.[330]

Ultimately, the Federal Circuit concluded that "a court's criticism of an attorney is simply commentary made in the course of an action to which the attorney is, legally speaking, a stranger."[331] Although the court recognized that "a formal reprimand constitutes a final decision . . . from which the sanctioned attorney may appeal," it qualified this statement by insisting that it "should not be taken to suggest that every statement criticizing an attorney or suggesting that the attorney has failed to comply with some legal or ethical norm amounts to a sanction sufficient to constitute a final decision in a collateral proceeding."[332] Essentially, the court refused to recognize the district court's commentary on Attorney Teschner's conduct as a final decision.[333] Therefore, Teschner had nothing to appeal.[334]

C. Petition for a Writ of Mandamus

By issuing a writ of mandamus, higher courts can compel lower courts or government officers "to perform mandatory or purely ministerial duties correctly."[335] This authority can be found in the All Writs Act of 1911, which permits courts to "issue all writs necessary or appropriate in aide of their respective jurisdictions and agreeable to the usages and principles of law."[336] The Supreme Court has prescribed that the:

---

327. Nisus Corp. v. Perma-Chink Sys., 479 F.3d 1316, 1318 (Fed. Cir. 2007).

328. *Id.* 28 U.S.C. § 1295 delineates the jurisdictional requirements for the Federal Circuit. The statute permits the Federal Circuit to hear appeals from U.S. district courts "if the jurisdiction of that court was based, in whole or in part, on section 1338 of this title." 28 U.S.C. § 1295(a)(1) (2006). Section 1338 states that "[t]he district courts shall have original jurisdiction of any civil action arising under any Act of Congress relating to patents, plant variety protection, copyrights and trademarks." 28 U.S.C. § 1338(a) (2006).

329. *Nisus*, 497 F.3d at 1318 (Fed. Cir. 2007) (citation omitted).

330. *Id.* at 1319–20

331. *Id.* at 1320.

332. *Id.*

333. *See id.* at 1320–21.

334. *See id.* at 1321.

335. BLACK'S LAW DICTIONARY 980 (8th ed. 2004) (defining the term "mandamus").

336. 28 U.S.C. § 1651(a) (2006).

> [T]raditional use of the writ in aid of appellate jurisdiction both at common law and in the federal courts has been to confine an inferior court to a lawful exercise of its prescribed jurisdiction or to compel it to exercise its authority when it is its duty to do so.[337]

Individuals who have been aggrieved by the district court and denied any opportunity to appeal may still petition the superior court for a writ of mandamus.[338] Indeed, the suggestion that mandamus would be a more appropriate remedy was made to the patent attorney in *Nisus Corp.*[339] Additionally, Attorney Pope's motion to intervene in *Therasense* was opposed using the same argument.[340]

However, a closer look at the doctrine of mandamus reveals that this is probably a dead end for patent attorneys. The writ of mandamus has been characterized as a "drastic remedy" and appropriate only in "extraordinary circumstances."[341] Anything short of a "judicial 'usurpation of power'" will likely be insufficient.[342] Furthermore, if the aggrieved party is able to obtain relief through appeal, then mandamus is inappropriate.[343] The Third Circuit has even gone so far as to say that the "writ is seldom issued and . . . discouraged."[344]

The petitioner for a writ of mandamus has the burden to establish that their right to relief is "clear and indisputable."[345] However, even if this requirement is satisfied and the petitioner has shown that there are no other avenues to seek relief, the issuance of mandamus is "largely discretionary."[346] Courts are reluctant to issue the writ largely due to the

---

337. Roche v. Evaporated Milk Assn., 319 U.S. 21, 26 (1943).

338. *See* 9 JAMES WILLIAM MOORE ET AL., MOORE'S FEDERAL PRACTICE ¶ 110.28 (2d ed. 1996) ("[T]he writ may issue to review conduct that is not otherwise reviewable by appeal.").

339. Nisus Corp. v. Perma-Chink Sys., Inc., 497 F.3d 1316, 1321–22 (Fed. Cir. 2007) (citation omitted).

340. Opposition to Pope's Motion for Leave to Intervene on Appeal at 10 (Bayer's Opposition Motion), Therasense, Inc. v. Becton, Dickinson & Co., 593 F.3d 1289 (Fed. Cir. 2008) (Nos. 2008-1511, -1512, -1513, -1514, -1595), *available at* http://www.patentlyo.com/bayer_s_20opposition_20to_20motion_20for_20leave_20to_20intervene_20by_20pope.pdf; Opposition to Motion for Leave to Intervene on Appeal at 20 (Becton, Dickinson & Co.'s Opposition Motion), Therasense, Inc. v. Becton, Dickinson & Co., 593 F.3d 1289 (Fed. Cir. 2008) (Nos. 2008-1511, -1512, -1513, -1514, -1595), *available at* http://www.patentlyo.com/bd_nova_s_20opposition_20to_20motion_20for_20leave_20to_20intervene_20on_20appeal_20re_20pope.pdf.

341. Hahnemann Univ. Hosp. v. Edgar, 74 F.3d 456, 461 (3d Cir. 1996).

342. *Id.* (quoting Will v. United States, 389 U.S. 90, 95 (1967)).

343. *Hahnemann*, 74 F.3d at 461 (citing Bankers Life & Casualty Co. v. Holland, 346 U.S. 379, 383 (1953)).

344. Lusardi v. Lechner, 855 F.2d 1062, 1070 (3d Cir. 1988).

345. *Bankers Life*, 346 U.S. at 384 (quoting United States v. Duell, 172 U.S. 576, 582 (1899)).

346. *Hahnemann*, 74 F.3d at 461.

effects of piecemeal appellate litigation on the "fair and prompt administration of justice."[347]

In light of the fact that district court findings of inequitable conduct are reviewed only for abuse of discretion,[348] it becomes difficult to imagine a scenario where an appellate court will ever issue a patent attorney's petition for mandamus. The attorney would have to prove that the district court clearly and indisputably abused its discretion in finding inequitable conduct.[349] Furthermore, superior courts would likely be hesitant to find the requisite extraordinary circumstances present because the patent owner is able to appeal the judgment itself and the patent attorney will have a hearing before the disciplinary board prior to any punishment.[350]

V. ANALYSIS

A. Why the Federal Circuit Got Things Wrong in Therasense

The Federal Circuit mistakenly relied too much on its earlier ruling in *Nisus Corp.* when it denied Attorney Pope's motion to intervene in *Therasense*.[351] The two situations were similar enough that, at first glance, the cases might have seemed analogous.[352] However, a deeper look at the circumstances of each case reveals that the Federal Circuit's reliance on *Nisus Corp.* in *Therasense* was misplaced.[353]

Unlike the original parties in *Nisus Corp.*, the original parties in *Therasense* were appealing the district court's decision.[354] In fact, the issue of inequitable conduct was slated for review by the Federal Circuit.[355] Therefore, rather than attempting to pursue an appeal on his own as Attorney Teschner was trying to do in *Nisus Corp.*, Attorney Pope was simply trying to join Abbot as an intervening party in *Therasense*.[356] The court's insistence that Pope should have been formally sanctioned by the trial court ignored the fact that the original parties were already appealing the issue of inequitable conduct.[357] Instead of analyzing the standard for the

347. Kerr v. United States Dist. Court, 426 U.S. 394, 403 (1976); *Hahnemann*, 74 F.3d at 461 (citing Haines v. Liggett Group Inc., 975 F.2d 81, 89 (3d Cir.1992)).

348. Kingsdown Med. Consultants v. Hollister, Inc., 863 F.2d 867, 876 (Fed. Cir. 1988).

349. *Id.* (quoting PPG Indus., Inc. v. Celanese Polymer Specialities Co., Inc., 840 F.2d 1565, 1572 (Fed. Cir. 1988)).

350. *Hahnemann*, 74 F.3d at 461 (citing *Haines*, 975 F.2d at 89).

351. *Therasense* Order, *supra* note 10.

352. Pope's Motion to Intervene, *supra* note 1.

353. *Therasense* Order, *supra* note 10.

354. *Id.*

355. *Id.*

356. *Id.*

357. *See* Pope's Reply in Support of His Motion for Leave to Intervene on Appeal, *supra* note 262.

appeal of sanctions,[358] the court should have considered the factors underlying intervention.

Even more surprising was the Federal Circuit's failure to analyze the standard for intervention under Ninth Circuit law.[359] The Federal Circuit was extremely clear in *Ericsson, Inc. v. Interdigital Communications Corp.* when it stated that a motion for intervention is "not unique to patent law . . . [and] therefore [is] governed by the law of the appropriate regional circuit."[360] Why then, did the Federal Circuit not analyze Attorney Pope's motion for intervention under Ninth Circuit law for this case that originated out of the Northern District of California?[361] By relying on *Nisus Corp.*, the Federal Circuit actually analyzed Attorney Pope's motion under Sixth Circuit law.[362]

One possible answer as to why the Federal Circuit erred in *Therasense* may be because of the way the court began its analysis in *Nisus Corp.*[363] Because the original parties were not pursuing an appeal, the court initially had to determine whether it had jurisdiction to hear Attorney Teschner's appeal,[364] noting that "[w]e resolve questions as to our jurisdiction by applying the law of this circuit, not the regional circuit from which the case arose."[365] The court lacked appellate jurisdiction largely because there had not been a final decision rendered on the issue that Teschner was attempting to appeal.[366]

This was in direct contrast to *Therasense*, where the court was not facing a jurisdictional issue on Attorney Pope's motion.[367] Jurisdiction had already been established when the original parties filed their Notice of Appeal.[368] In light of the Supreme Court's ruling in *McConnell*[369] and the

358. *Therasense* Order, *supra* note 10.

359. *Id.*

360. Ericsson, Inc. v. Interdigital Commc'n. Corp., 418 F.3d 1217, 1221 (Fed. Cir. 2005).

361. Therasense, Inc. v. Becton, Dickinson & Co., 565 F. Supp. 2d 1088 (N.D. Cal. 2008).

362. *See* Nisus Corp. v. Perma-Chink Sys., Inc., 497 F.3d 1316, 1323 n.1 (Fed. Cir. 2007) ("[T]here is no reason to believe Sixth Circuit law would permit Mr. Teschner to intervene for purposes of pursing an appeal in the present case under these circumstances.").

363. *Nisus Corp.*, 497 F.3d at 1318–19 (citing Silicon Image, Inc. v. Genesis Microchip, Inc., 395 F.3d 1358, 1362 (Fed. Cir. 2005)).

364. *Id.*

365. *Id.*

366. *Nisus Corp.*, 497 F.3d at 1320.

367. Pope's Motion to Intervene, *supra* note 1.

368. *Id.*

369. McConnell v. FEC, 540 U.S. 93, 233 (2003) ("It is clear, however, that the Federal Election Commission (FEC) has standing, and therefore we need not address the standing of the intervenor-defendants, whose position here is identical to the FEC's." (citing Bowsher v. Synar, 478 U.S. 713, 721 (1986); Diamond v. Charles, 476 U.S. 54, 68–69, n.21 (1986))).

Ninth Circuit's ruling in *Portland Audubon Society*,[370] the Federal Circuit should have granted Attorney Pope's motion to intervene.

Furthermore, the jurisdictional question of appealing judicial findings or sanctions from the trial court is not unique to patent law,[371] as evidenced by the multiple decisions on the issue from nearly every circuit. Therefore, instead of using Federal Circuit law, the law of the regional circuit would be more appropriate.[372] Had this been the standard for *Nisus Corp.*, the outcome would still have been uncertain.[373] The Sixth Circuit's opinion in *Associated Builders & Contractors v. Perry* hints that Attorney Teschner's motion would still have been denied.[374] On the other hand, the Ninth Circuit has specifically designated that findings of an attorney's ethical rule violation are per se sanctions that can be appealed.[375] Therefore, under Ninth Circuit law,[376] Attorney Pope should also have been permitted to appeal the district court's findings against him.

B. Patent Attorneys Satisfy the Requirements for Permissive Intervention

The standard for permissive intervention, expressed in Rule 24(b), is satisfied upon a patent attorney's timely application to intervene.[377] The patent attorney will always be assuming a defensive posture when intervening to *defend* against claims of inequitable conduct.[378] Because the patent owner in the case will have to defend against the inequitable conduct charges as well, common questions of law and fact are inevitable.[379] The Supreme Court has also clarified that standing is not required for an intervenor.[380] One commentator has written that "[f]or intervenors merely protecting their interests in a defensive posture or seeking to bring claims

---

370. Portland Audubon Soc'y v. Hodel, 866 F.2d 302, 308, n.1 (9th Cir. 1989) (noting that, historically, they "have resolved intervention questions without making reference to standing doctrine").

371. Ericsson, Inc. v. Interdigital Commc'n. Corp., 418 F.3d 1217, 1220–21 (Fed. Cir. 2005).

372. Nisus Corp. v. Perma-Chink Sys., Inc., 497 F.3d, 1316, 1318–19 (Fed. Cir. 2007) (citing Silicon Image, Inc. v. Genesis Microchip, Inc., 395 F.3d 1358, 1362 (Fed. Cir. 2005)).

373. *Id.*

374. 16 F.3d 688, 692 (6th Cir. 1994) (holding that NECA lacked standing to appeal because the State of Michigan (the original party) declined to prosecute the appeal).

375. United States v. Talao, 222 F.3d 1133, 1138 (9th Cir. 2000).

376. *Id.*

377. Kootenai v. Veneman, 313 F.3d 1094, 1111 (9th Cir. 2002).

378. *Id.*

379. *Id.*

380. McConnell v. FEC, 540 U.S. 93, 233 (2003) ("It is clear, however, that the Federal Election Commission (FEC) has standing, and therefore we need not address the standing of the intervenor-defendants, whose position here is identical to the FEC's.") (citing Bowsher v. Synar, 478 U.S. 714, 721 (1986); Diamond v. Charles, 476 U.S. 54, 68–69, n.21 (1986)).

identical to those already presented by the original parties, the intervenors should not be required to show an interest in the litigation rising to the level of standing."[381] Under this logic, the patent attorney would be seeking to intervene on the side of the patent owner for the purpose of protecting the patent attorney's own professional reputation.[382]

Not only will an intervening patent attorney satisfy the requirements for permissive intervention, but the accused infringer will not be prejudiced.[383] First, the accused infringer raised the claims of inequitable conduct as an affirmative defense to the alleged patent infringement.[384] Second, the original party, usually a patent owner, will automatically defend against the claims.[385] As long as the patent attorney requests no additional time, then prejudice will not be an issue.[386]

C. A Statutory Right of Intervention Is Necessary

One of the purposes of this Comment is to ask whether patent attorneys faced with claims of inequitable conduct have sufficient recourse to defend their professional reputations. In multiple circuits there are legitimate questions as to whether patent attorneys have *any* opportunity to participate in their defense until the PTO initiates a disciplinary proceeding.[387] The circuit confusion is particularly frustrating when it seems that the requirements for permissive intervention have been satisfied by the accused patent attorney.[388] The law revolving around the appealability of attorney sanctions is even more murky, and it is hard to justify calling the writ of mandamus a recourse opportunity. A statutory right of intervention is needed for patent attorneys when claims of inequitable conduct have been lodged against them and those claims are still being litigated, including on appeal, by the original parties.

A statutory right of intervention would upgrade the current regime. A statute would allow patent attorneys to circumvent the conflicts that exist among the circuits over the application of intervention in their unique context. This would also be the best way to advance the notion of uniform federal law for patents and improve predictability and consistency in patent litigation.

Participation at trial would also provide the PTO with more information upon which to judge the attorney's conduct when it considers any

381. *See* Timmermans, *supra* note 169, at 1441.
382. *See id.*
383. *See id.*
384. *See id.*
385. *See id.*
386. *See id.*
387. Nisus Corp. v. Perma-Chink Sys., Inc., 497 F.3d 1316, 1320 (Fed. Cir. 2007).
388. Pope's Motion to Intervene, *supra* note 1.

disciplinary action against the attorney.[389] The involvement of the patent attorney in the case would "contribute to the equitable resolution of [the] case."[390] At the very least, it would be more equitable for the patent attorney faced with the charges.[391] Participation before a court has "crystallized its views"[392] allows the patent attorney an opportunity to argue in front of an unbiased tribunal. On the other hand, forcing the attorney to wait and champion the attorney's cause until the PTO has commenced a disciplinary action will mean that the defense will always follow a district court opinion and be ineffective against the practical effects of a published opinion.

Currently, there are twenty-three federal statutes authorizing parties to intervene as of right.[393] All twenty-three share language that says the party or individual "may intervene as a matter of right."[394] Some statutes place conditions on the statutory right[395] while others provide for an absolute right.[396] The proper statutory right of intervention for patent attorneys should limit the right only to situations where the original plaintiffs are still pursuing the litigation and the attorney's inequitable conduct is still an issue.

Some critics have opposed the notion of allowing patent attorneys to *appeal* a finding of inequitable conduct.[397] Some of these same arguments[398] may also be used to oppose a statutory right of intervention for patent attorneys accused of inequitable conduct. However, permitting only intervention while the original parties are still litigating the issue would

---

389. Kootenai Tribe of Idaho v. Veneman, 313 F.3d 1094, 1111 (9th Cir. 2002).

390. *Id.*

391. *Id.*

392. Int'l Union v. Scofield (*Auto Workers*), 382 U.S. 205, 213 (1965).

393. 15 U.S.C. § 2619(c) (2006); 15 U.S.C. § 3205 (2006); 15 U.S.C. § 3371(b)(2)(D) (2006); 15 U.S.C. § 3391(d) (2006); 15 U.S.C. § 3415 (2006); 16 U.S.C. § 544m(b)(3)(A) (2006); 21 U.S.C. § 337 (2006); 28 U.S.C. § 2631(j)–(k) (2006); 30 U.S.C. § 1270(c) (2006); 33 U.S.C. § 1365(c)(2) (2006); 33 U.S.C. § 1515(b)–(c) (2006); 42 U.S.C. §300j-8(c) (2006); 42 U.S.C. § 2000e-5(f) (2006); 42 U.S.C. § 4911(c) (2006); 42 U.S.C. § 6305(c) (2006); 42 U.S.C. § 6972(c) (2006); 42 U.S.C. § 7604(c) (2006); 42 U.S.C. § 8435(c) (2006); 42 U.S.C. § 9124(c) (2006); 42 U.S.C. § 9613(i) (2006); 42 U.S.C. § 9659(g) (2006); 42 U.S.C. § 11046(h) (2006); 43 U.S.C. § 1349(a)(4) (2006).

394. *Id.*

395. *See e.g.*, § 2619(b)(1) (stating that once the Administrator has commenced a proceeding, only persons that gave notice prior to the Administrator's commencement may intervene as of right).

396. *See e.g.*, § 2619(c)(1) (stating that the Administrator, "if not a party, may intervene as a matter of right.").

397. *See* Nisus Corp. v. Perma-Chink Sys., Inc., 497 F.3d 1316, 1320 (Fed. Cir. 2007); Bolte v. Home Ins. Co., 744 F.2d 572, 573 (1984); Matthew Funk, *Sticks and Stones: The Ability of Attorneys to Appeal From Judicial Criticism*, 157 U. PA. L. REV. 1485, 1508–09 (2009).

398. *See* Funk, *supra* note 397.

avoid nearly all of the concerns cited in their arguments. For instance, the Federal Circuit in *Nisus Corp.* warned that allowing appeals by attorneys "concerned about their professional or public reputations merely because a court criticized them . . . would stretch the concept of collateral proceedings . . . [and] result in a multiplicity of appeals."[399] However, as the dissent points out in *Williams v. United States*:

> [D]efining a certain limited class of reprimands as appealable under 28 U.S.C. § 1291 and § 158(d) is unlikely to prompt a flood of new appeals to this already busy court. Any lawyer appealing a reprimand takes the risk that this court, reaching the merits, will agree that the sanction is justified, thus giving the sanction far more force than it would have had if it had come from a trial judge unendorsed by a reviewing court. Accordingly, the lawyer's self-interest dictates that an appeal be taken only in cases in which the sanction is particularly damaging to the lawyer's reputation and particularly undeserved.[400]

Thus, allowing a patent attorney accused of inequitable conduct to intervene in an active litigation would subject the attorney to the risk that the reviewing court will agree that the sanction was justified.[401] Moreover, an inequitable conduct finding will nearly always be "particularly damaging to the lawyer's reputation."[402] Finally, limiting intervention exclusively to active litigation dispenses with any concern of "a breathtaking expansion in appellate [litigation.]"[403]

Additionally, one scholar, Matthew Funk, is critical of patent attorneys being allowed to appeal the district court findings because their conduct did not occur during the litigation.[404] While insisting that the justifications for allowing only appeals by attorneys are "weakened" in this situation, the only reason Funk cites is the difficulty in justifying why "witnesses or third parties who have findings of misconduct against them" could not appeal.[405] This same logic[406] could also be used to argue against a special statutory right of intervention for patent attorneys facing allegations of inequitable conduct. In response, inequitable conduct and patent attorneys, like the rest of the patent regime, are entirely unique to patent law. Unlike witnesses or third parties who have been accused of misconduct during the trial, the patent attorney's conduct is the focus of the litigation on the issue of inequitable conduct[407] and not some peripheral concern.

---

399. *Nisus Corp.*, 497 F.3d at 1320.
400. 158 F.3d 50, 51 (1st Cir. 1998) (Lynch, C.J., dissenting).
401. *Id.*
402. *Id.*
403. Bolte v. Home Ins. Co., 744 F.2d 572, 573 (1984).
404. *See* Funk, *supra* note 397, at 1508–09.
405. *Id.*
406. *See id.*
407. Therasense, Inc. v. Becton, Dickinson & Co., 593 F.3d 1289, 1300–11 (Fed. Cir.

A finding of inequitable conduct against a patent attorney is much more debilitating to an attorney's career than a finding of misconduct against a witness or even a trial attorney.[408] Aside from any general consequences to legal problems, it is unlikely that lay witnesses professional careers will be affected by litigation to which they are not a party. A sanctioned trial attorney could likely characterize their reprimand as vigor for their clients.

Conversely, clients will lose confidence in the patent attorney once the attorney has been found to have committed inequitable conduct. The perceived strength of any patent that attorney had already prosecuted would surely be weakened. The goal of obtaining a patent is to be able to enforce it without litigation. Accused infringers will be more likely to forego licensing, infringe, and proceed with litigation if they believe that they can easily defeat the charges using the inequitable conduct defense. Proving unenforceability in litigation could be less than paying the patent owner's licensing fee.

VI. CONCLUSION

Today, the success of a patent attorney's attempt to intervene in litigation where the attorney has been accused of inequitable conduct is largely circuit dependent. Relying on the jurisprudence of appealing sanctions offers even less promise to the accused patent attorney. Courts that have denied both of these avenues point to the writ of mandamus as the proper forum for patent attorneys to address a district court's findings of inequitable conduct against them.[409] Unfortunately, that too is a dead end.[410] As a result, this area of patent law is unpredictable and inconsistent.

The effects of an inequitable conduct finding on a patent attorney's career are undisputed. Aside from the possibility of losing certification to practice in front of the PTO,[411] a finding of inequitable conduct will be debilitating on the attorneys professional reputation. Indeed, Attorney Pope not only received inquiries from disciplinary bodies after the district court's opinion was rendered but was also the subject of two magazine articles.[412]

2010).

408. An attorney's conduct in the course of trial may be interpreted as zealous advocacy for their clients. An inequitable conduct finding, however, brings into question the reliability and validity of all past and future patents that the attorney prosecuted. *See* Elinson, *supra* note 13; *see also supra* text accompanying note 125.

409. Nisus Corp. v. Perma-Chink Sys., Inc., 497 F.3d 1316, 1321–22 (Fed. Cir. 2007).

410. Lusardi v. Lechner, 855 F.2d 1062, 1069 (3d. Cir. 1988) (characterizing the writ of mandamus as "seldom issued and . . . discouraged").

411. 37 C.F.R. §§ 11.20 (2009) (PTO discipline can include exclusion, suspension, reprimand, censure, or probation).

412. *See* Elinson, *supra* note 13; Miller, *supra* note 13, at 17–18.

The Federal Circuit was wrong in *Therasense* when it refused to allow Attorney Pope to intervene.[413] Because the concept of intervention is not unique to patent law,[414] the Federal Circuit should have analyzed Attorney Pope's Motion to Intervene under Ninth Circuit law. Its prior holding in *Nisus Corp.*[415] should not have controlled because *Nisus Corp* originated in the Sixth Circuit[416] and, unlike *Nisus Corp.*, the original parties in *Therasense* were still pursuing the inequitable conduct issue on appeal.

In sum, a statutory right of intervention is needed for patent attorneys faced with inequitable conduct allegations as long as the litigation is ongoing and the original parties are still disputing that issue. A statutory right will increase consistency and uniformity in the patent regime. Participation on the part of the patent attorney will "contribute to the equitable resolution of [the] case"[417] and provide the PTO with more information upon which to base their disciplinary decisions.

413. *Therasense* Order, *supra* note 10.

414. Ericsson, Inc. v. Interdigital Commc'n. Corp., 418 F.3d 1217, 1220–21 (Fed. Cir. 2005).

415. Nisus Corp. v. Perma-Chink Sys., Inc., 497 F.3d 1316, 1322 (Fed. Cir. 2007).

416. Nisus Corp v. Perma-Chink Sys., Inc., No. 3:03-CV-120, 2006 WL 2128903 (E.D. Tenn. 2006).

417. Kootenai Tribe of Idaho v. Veneman, 313 F.3d 1094,1111 (9th Cir. 2002).

THE RELIGIOUS LAND USE AND INSTITUTIONALIZED PERSONS ACT: A PERSPECTIVE ON THE UNREASONABLE LIMITATIONS PROVISION

YUSUF Z. MALIK*

I. INTRODUCTION 532
A. *Background* 537
B. *RLUIPA: Legislative Findings* 540
II. 42 U.S.C. § 2000CC(B)(3)(B)—THE UNREASONABLE LIMITATIONS PROVISION 542
A. *Legislative History of Unreasonable Limits* 543
B. *Case Law: The Unreasonable Limitations Provision* 545
C. RMCC II *and its Importance for RLUIPA* 546
III. RLUIPA'S UNREASONABLE LIMITATIONS PROVISION AND CASES ANALYZED 551
A. *An Analysis of* RMCC II's *Outcome with Regard to the Unreasonable Limitations Provision* 551
B. *Boulder County's Land Use Policy and Rationale for Excluding Churches* 552
C. *Unreasonable Limitations:* RMCC II *in Light of Other Cases* 554
1. Unreasonable Limitations and Expansion 554
2. Expansion Opportunities Compared under the Substantial Burden Provision versus Unreasonable Limitations Provision 556
3. Unreasonable Limitations and Undue Discretion 558
IV. FINDING COMMON GROUND 560
A. *Land Use Regulation or Religious Freedom?* 560
B. *A Future Perspective* 562
V. CONCLUSION 567

I. INTRODUCTION

The Religious Land Use and Institutionalized Persons Act ("RLUIPA")[1], a federal statute, has been the subject of ongoing controversy since its passage in 2000.[2] In the land use context, RLUIPA prohibits a state or local government[3] from imposing or implementing a land use regulation[4] that substantially burdens the "religious exercise of a person, including a religious assembly or institution . . . ."[5] To overcome this prohibition, a local government needs to show that the law enacted both furthers a compelling governmental interest and is narrowly tailored to achieve that interest.[6] This standard of review is referred to as the strict scrutiny test, which is a difficult burden for a government to overcome.[7]

---

* Candidate for Doctor of Jurisprudence, May 2011, University of Tennessee College of Law. I would like to thank Professor Gregory Stein who gave me suggestions for this Comment and provided me with guidance.

1. Religious Land Use and Institutionalized Persons Act of 2000, 42 U.S.C. § 2000cc (2006). This Comment will not address the Institutionalized Persons portion of the Act, which was ruled constitutional in *Cutter v. Wilkinson*, 544 U.S. 709, 709 (2005).

2. *See infra* Part I.A.

3. A "government" defined under RLUIPA means: "(i) a State, county, municipality or other government entity created under the authority of a State; (ii) any branch, department, agency, instrumentality, or official of an entity listed in clause (i); and (iii) any other person acting under color of State law . . . ." 42 U.S.C. § 2000cc-5(4) (2006).

4. "Land use regulation" defined under RLUIPA means:

> [A] zoning or landmarking law, or the application of such a law, that limits or restricts a claimant's use or development of land (including a structure affixed to land), if the claimant has an ownership, leasehold, easement, servitude, or other property interest in the regulated land or a contract or option to acquire such an interest.

42 U.S.C. § 2000cc-5(5) (2006).

5. 42 U.S.C. § 2000cc(a)(1).

6. 42 U.S.C. § 2000cc(a). This section represents the substantial burden portion of RLUIPA and provides:

> (1) No government shall impose or implement a land use regulation in a manner that imposes a substantial burden on the religious exercise of a person, including a religious assembly or institution, unless the government demonstrates that the imposition of the burden on that person, assembly, or institution—
> (A) is in furtherance of a compelling governmental interest; and
> (B) is the least restrictive means of furthering that compelling governmental interest.

Id. Thus far, what it means to "substantially burden" a person's religious exercise has been construed differently among circuits. *Compare* Civil Liberties for Urban Believers v. City of

In addition to the substantial burden category, the land use portion of RLUIPA includes a discrimination and exclusions category.[8] The unreasonable limitations provision, which is within that category, states that "[n]o government shall impose or implement a land use regulation that . . . unreasonably limits religious assemblies, institutions, or structures within a jurisdiction."[9] Before discussing this particular provision in depth, it is important to understand how these RLUIPA categories operate.

Consider the recent controversy in New York City involving the building of a mosque near the site of the 9/11 terrorist attacks. The building developers cleared their last hurdle after the city's Landmarks Preservation Commission voted against preserving the current structure on the site where the mosque will be built.[10] However, vehement community opposition remains, particularly from those who had family members killed on 9/11.[11] RLUIPA "lurks in the background" in this circumstance because its land use provisions would provide grounds for a federal lawsuit if the land use authorities involved in this decision denied the mosque the right to build.[12] Specifically, based on RLUIPA's substantial burden category, the mosque leaders would have grounds to allege that the government has placed a substantial burden on its religious exercise by denying it the right to build.[13]

The equal terms section, embedded within RLUIPA's discrimination and exclusions category, could also apply to the so-called "Ground Zero

---

Chicago, 342 F.3d 752, 761 (7th Cir. 2003) (holding that a zoning ordinance did not impose a substantial burden on religious exercise based on an impracticality standard) *with* San Jose Christian Coll. v. City of Morgan Hill, 360 F.3d 1024, 1034–35 (9th Cir. 2004) (holding that a denial of a re-zoning application did not substantially burden a college's religious exercise because the regulation at issue was not "oppressive" to a "significantly great" extent).

For an exposition of these cases and the various standards followed by courts for construing substantial burden, see Adam J. MacLeod, *A Non-Fatal Collision: Interpreting RLUIPA Where Religious Land Uses and Community Interests Meet*, 42 URB. LAW. 41 (2010). This Comment will touch on the substantial burden provision and its relation to the unreasonable limitations provision within the context of expansion.

7. *See* Rocky Mountain Christian Church v. Bd. of Cnty. Comm'rs of Boulder Cnty., *(RMCC I)* 612 F. Supp. 2d 1163, 1174–75 (D. Colo. 2009) (holding that the interests served by Boulder County's Land Use Code and the County's Comprehensive Plan did not constitute compelling governmental interests).

8. *See* 42 U.S.C. § 2000cc(b) (2006); *see also infra* note 80 and accompanying text (providing the full discrimination and exclusions category).

9. 42 U.S.C. § 2000cc(b)(3)(B) (2006).

10. Javier C. Hernandez, *Mosque Near Ground Zero Clears Key Hurdle*, CITY ROOM BLOG–N.Y. TIMES (Aug. 3, 2010, 10:38 AM), http://cityroom.blogs.nytimes.com/2010/08/03/mosque-near-ground-zero-clears-key-hurdle/.

11. *Id.*

12. Marci A. Hamilton, *The Wrong-Headed Furor over the Planned Mosque at Ground Zero: Mistaking a War on Radical Islamicsim for a War on All Muslims*, FINDLAW (Aug. 5, 2010), http://writ.news.findlaw.com/hamilton/20100805.html.

13. *See supra* note 6.

mosque" because that section prevents the government from "treat[ing] a religious assembly or institution on less than equal terms with a nonreligious assembly or institution."[14] For instance, the mosque developers might have been on less equal terms if businesses were allowed to build near the 9/11 grounds, but they were not. Further, if the land use authorities involved in the mosque decision engaged in past restrictive practices on religious users in the land use context, the mosque's leaders would have grounds for an unreasonable limitations claim.[15] The group could allege, in accord with that provision, that the government had unreasonably limited religious institutions within a jurisdiction.[16] Thus, RLUIPA's land use provisions can provide a religious land user with various legal avenues when its religious exercise is infringed upon.

The language of RLUIPA's unreasonable limitations provision seems relatively straightforward. However, its application in cases raises questions such as what it means to "unreasonably limit."[17] RLUIPA does not define "unreasonably limits," and the absence of such a definition illustrates what has become a common problem with RLUIPA—its broad and vague statutory language.[18] The absence of an unreasonable limitations definition within RLUIPA places a burden on municipalities, religious claimants, and courts.

Municipalities are not given guidance by the provision's language on what land use regulations might be unreasonable. As a result, land use planning can become complicated, especially when a municipality's intended use for a particular plot of land is prone to a religious entity's takeover by using the provision. Religious claimants can—and do—take over land based on claims that they have been unreasonably restricted from expanding or locating in a particular jurisdiction.[19]

The religious claimant, although in a less precarious position than a municipality, is also burdened by the absence of definition. The provision's language affects how and if a claimant will plead a RLUIPA claim. Since the provision is not clear as to what constitutes an unreasonable limitation,

---

14. 42 U.S.C. § 2000cc(b)(1) ("No government shall impose or implement a land use regulation in a manner that treats a religious assembly or institution on less than equal terms with a nonreligious assembly or institution.").

15. *See* 42 U.S.C. § 2000cc(b)(3)(B).

16. *Id.*

17. *See, e.g.*, Guru Nanak Sikh Soc'y of Yuba City v. Cnty. of Sutter, 326 F. Supp. 2d 1140, 1155 (E.D. Cal. 2003) (noting that the application of some of RLUIPA's provisions will certainly encounter difficulties in the future when interpreting the statute).

18. *See* MacLeod, *supra* note 6, at 44–62. "Land use regulation," "religious exercise," and "substantial burden" are subject to various interpretations. *See id.*

19. *See e.g.*, Rocky Mountain Christian Church v. Bd. of Cnty. Comm'rs of Boulder Cnty., (*RMCC II*) 613 F.3d 1229, 1239–40 (10th Cir. 2010).

religious claimants may be unsure if a violation of the provision has occurred.[20]

Finally, courts are potentially burdened by RLUIPA's lack of definition for "unreasonably limits." More plaintiffs can use this lack of a definition to construct an argument against a local government in their favor, resulting in more claims. More claims means more appeals. Further, more claims clog courts' dockets, a clogging that would not have occurred but for the vague and ambiguous drafting of RLUIPA. The uncertainty about the unreasonable limitations provision is further exacerbated by the minimal yet growing case law in existence addressing the issue.[21] Lawsuits arise when religious freedom is favored over land use standards, or vice versa. Unless courts produce a clear standard on what constitutes an unreasonable limitation, claimants are left with bits of RLUIPA's legislative history and case precedent to make sense of the unreasonable limitations provision.

First, this Comment will briefly recount the history of RLUIPA.[22] Part II will address the terms of the unreasonable limitations provision, its legislative history, and applicable cases including the unreasonable limitations provision in light of the 2010 Tenth Circuit case *Rocky Mountain Christian Church v. Board of County Commissioners of Boulder County*.[23] Part III will analyze the unreasonable limitations under *RMCC II*, discuss Boulder County's land use policy, and its possible rationales for restricting church growth. Also, Part III of the Comment will address the unreasonable limtations provision's application when a religious claimant is trying to expand its facilities.

---

20. Adhi Parasakthi Charitable, Med., Educ., & Cultural Soc'y of N. Am. v. Twp. of W. Pikeland, (*West Pikeland*) No. 09-cv-1626, 2010 WL 2635979, at *17 (E.D. Pa. June 25, 2010) (Religious claimant alleging a violation of the provision based on a "limitation on the size of a building on its property to 5,000 square feet."); *see* Grace Church of Roaring Fork Valley v. Bd. of Cnty. Comm'rs of Pitkin Cnty., (*Grace Church*) No. 05-cv-01673-RPM, 2010 WL 3777286, at *8 (D. Colo. Sept. 20, 2010) (holding that the plaintiffs' unreasonable limitations claim failed because the high cost of real estate, which impeded new church development, was not evidence of a violation of the provision).

21. Most of the significant cases, which will be covered in this Comment, include: *RMCC II*, 613 F.3d at 1229; Vision Church, United Methodist v. Vill. of Long Grove, (*Vision Church*) 468 F.3d 975, 990–91 (7th Cir. 2006); *West Pikeland*, 2010 WL 2635979, at * 17; Chabad of Nova, Inc. v. City of Cooper City, 575 F. Supp. 2d 1280 (S.D. Fla. 2008); Konikov v. Orange Cnty., Florida, 302 F. Supp. 2d 1328, 1345–46 (M.D. Fla. 2004), *aff'd in part, rev'd in part*, 410 F.3d 1317 (11th Cir. 2005). The case law is starting to expand on RLUIPA's unreasonable limitations provision. *See, e.g.*, *Grace Church*, 2010 WL 3777286, at *3 (plaintiff church alleged a violation of the provision because their special review application was denied).

22. There has already been much existing literature on this topic. *See, e.g.*, Daniel P. Lennington, *Thou Shalt Not Zone: The Overbroad Applications and Troubling Implications of RLUIPA's Land Use Provisions*, 29 SEATTLE U. L. REV. 805, 811–19 (2006).

23. *RMCC II*, 613 F.3d 1229.

This Comment asserts that while the unreasonable limitations provision has a limited application, a land use authority should still be wary of the provision as the *RMCC II* case demonstrates.[24] A religious claimant seeking to expand its facilities can trigger the provision by showing that a municipality unnecessarily created a restrictive environment for religious users in the land use context.[25] Part III concludes by assessing whether a local land use authority's overly broad discretion in the land use context violates RLUIPA's unreasonable limitations provision.

Part IV of the Comment will assess the implications of the unreasonable limitations provision and RLUIPA on land use regulations, particularly their potential effect on comprehensive plans.[26] The unreasonable limitations provision and RLUIPA give religious claimants too much ability to frustrate land use goals, causing potential distress to communities. The broad religious protection afforded under RLUIPA creates an environment where governments potentially adopt more lenient land use standards because of the prospect of a RLUIPA lawsuit.[27] Flexible land use standards are not always desirable, especially for a city that does not have the necessary infrastructure to support religious building growth. However, RLUIPA *is* needed to resolve restrictive practices in the land use context.[28] If a religious entity has been limited from a jurisdiction, it should have its recourse. This is what RLUIPA, and more specifically, RLUIPA's unreasonable limitations provision, provides.

With these competing interests in mind, Part IV will attempt to bridge the rift between land use authorities and religious institutions. Also, in light

---

24. *Id.*

25. *See infra* Part III.C.1.

26. Comprehensive plans serve as guidance for area development and proposed future land use and zoning. 83 AM. JUR. 2D *Zoning and Planning* § 17 (2003). *See generally* Daniel J. Curtin, Jr., *Ramapo's Impact on the Comprehensive Plan*, 35 URB. LAW. 135, 136 (2003) ("Increasingly, local jurisdictions are implementing the General Plan [i.e. Comprehensive Plan] as part of their land use planning process. Although specifics vary widely, most jurisdictions with a General Plan view it as the 'constitution' for development within that community. Typically, all subsequent land use decisions must be 'consistent' with the vision for growth and development reflected in the General Plan."). *Id.*; *see also* ROBERT C. ELLICKSON & VICKI L. BEEN, LAND USE CONTROLS: CASES AND MATERIALS 74 (3d ed. 2005) ("Comprehensive Plans consist principally of (a) statement of goals and (b) maps that establish use and density guidelines for various districts and project future public improvements.").

27. *See* Heather M. Welch, *The Religious Land Use and Institutionalized Persons Act and Mega-Churches: Demonstrating the Limits of Religious Land Use Exemptions in Federal Legislation*, 39 U. BALT. L. REV. 255, 260 (2009).

28. 146 CONG. REC. 16,698, 16,699 (2000) (joint statement of Sens. Hatch and Kennedy) ("[D]iscrimination against religious uses is a nationwide problem. . . . Where it occurs, it is often covert. It is impossible to make separate findings about every jurisdiction, or to legislate in a way that reaches only those jurisdictions that are guilty.").

of the unreasonable limitations provision's literal language,[29] Part IV will touch on the absence of certain limitations that exist on larger religious institutions, focusing on the proliferation of so-called megachurches.[30] In retrospect, RLUIPA's drafters did not necessarily anticipate the growth nor the protection of megachurches and their attendant secular amenities. These types of churches and their multipurpose facilities[31] expose a certain inadequacy in RLUIPA in that the statute fails to address secular land use by religious institutions.[32]

Part IV concludes that this particular megachurch loophole, coupled with RLUIPA's vague provisions, such as the substantial burden and the unreasonable limitations provisions, exposes the need for federal legislation to help cure future land use planning defects. Finally, Part V will provide a brief conclusion.

A. Background

In 2000, Congress passed RLUIPA to address the limitations local governments sometimes place on the free exercise of religion in the context of land use and institutionalized persons.[33] President Clinton signed

---

29. The provision brings about the idea of expansion and limitations by its language: "No government shall impose or implement a land use regulation that . . . unreasonably limits religious assemblies, institutions, or structures within a jurisdiction." 42 U.S.C. § 2000cc(b)(3)(B) (2006).

30. The limitations on so-called "mega-churches" will become an important issue for the future. *See generally* Welch, *supra* note 27, at 256 (noting that RLUIPA is "ill equipped" to address the secular land use problem posed by megachurches).

31. *Id.* ("A mega-church is not only a house for services and prayer, but it is also a one-stop shop for congregants—an all inclusive community where people can 'eat, shop, go to school, bank, work out, scale a rock-climbing wall and pray . . . all without leaving the grounds.'") (quoting Patricia Leigh Brown, *Megachurches as Minitowns*, N.Y. TIMES, May 9, 2002, at F1). In addition to megachurches, other churches are using land more intensively. *See generally* Alan C. Weinstein, *How to Avoid a "Holy War"—Dealing with Potential RLUIPA Claims*, 60 PLAN. & ENVTL. L. 3, 4 (March 2008) ("[M]any churches sponsor a school, day care center, adult education classes, a variety of programming serving different age groups, and various faith-based 'support' groups.").

32. *See* Welch, *supra* note 27, at 256.

33. 146 CONG. REC. 16,698 (2000) (joint statement of Sens. Hatch and Kennedy). Generally, the traditional state court view was to strike down ordinances "that excluded houses of worship from residential districts." ELLICKSON & BEEN, *supra* note 26, at 209. "Those courts sometimes based their decision on the notion that religious uses of land are 'inherently beneficial,' and so obviously further the public health, safety, and welfare that attempts to bar them from residential districts are irrational." *Id.* (citing Diocese of Rochester v. Planning Bd., 136 N.E.2d 827 (N.Y. 1956)). Lower federal court decisions digressed from this view in the early 1980s. *Id.*; *see, e.g.*, Lakewood, Ohio Congregation of Jehovah's Witnesses, Inc. v. City of Lakewood, 699 F.2d 303, 308 (6th Cir. 1983) (holding that a zoning ordinance, which restricted a congregation from building a house of worship on a purchased lot, did not violate the due process clause).

RLUIPA into law on September 22, 2000.[34] As will be explained in greater detail, the statute's passage was the result of a ten-year conflict between Congress and the Supreme Court over religious protection. The 1990 Supreme Court decision in *Employment Division v. Smith* marked the advent of this struggle, where the statute's origins can be traced.[35]

In *Smith*, the respondents, Native American peyote users, challenged the constitutionality of an Oregon statute that criminalized possession of peyote.[36] Oregon had denied their petition for unemployment benefits because the respondents were dismissed from their jobs for "misconduct" for the ingestion of peyote.[37] In *Smith*, the issue was whether the prohibition of peyote use was permissible under the Free Exercise Clause of the First Amendment, as the respondents used the peyote for religious purposes.[38] The Supreme Court reasoned that the law outlawing peyote was neutral and of general applicability so as not to implicate the Free Exercise Clause.[39] As a result, the Court upheld the constitutionality of the Oregon state law prohibiting possession of peyote.[40]

Smith had the effect of limiting the strict scrutiny test employed in the 1963 decision *Sherbert v. Verner*, where the Supreme Court reasoned that no government could enforce a law in a manner that burdened the exercise of a person's religion unless a compelling interest was established through the least restrictive means available.[41] *Smith* limited *Sherbert* by employing a rational basis test for generally applicable state laws.[42] The case also limited free exercise protection for religious conduct to circumstances where a law is specifically aimed toward religious practice.[43] Similarly, *Smith* undermined the ruling of *Wisconsin v. Yoder*. [44] In *Yoder*, the

---

34. *See* Statement on the Signing of the Religious Land Use and Institutionalized Persons Act of 2000, 2 PUB. PAPERS 1905, 1906 (Sept. 22, 2000).

35. Emp't Div., Dep't of Human Res. of Or. v. Smith, 494 U.S. 872, 890 (1990), *superseded by statute*, Religious Freedom Restoration Act (RFRA) of 1993, PUB. L. NO. 103-141, 107 Stat. 1488 (invalidated as applied to the states by City of Boerne v. Flores, 521 U.S. 507 (1997)); *see* Madison v. Riter, 355 F.3d 310, 314 (4th Cir. 2003) (noting that Congress modeled RLUIPA to conform to the decisions in *Smith* and *Flores*).

36. *Smith*, 494 U.S. at 878.

37. *Id.* at 874.

38. *Id.* at 876.

39. *Id.* at 878–82.

40. *Id.* at 890.

41. Sherbert v. Verner, 374 U.S. 398, 403 (1963).

42. *Smith*, 494 U.S. at 884–85.

43. ALLAN IDES & CHRISTOPHER N. MAY, CONSTITUTIONAL LAW INDIVIDUAL RIGHTS 495–96 (5th ed. 2010). The *Smith* Court reasoned that religious freedom should receive a great degree of protection, stating that "[t]he free exercise of religion means, first and foremost, the right to believe and profess whatever religious doctrine one desires. Thus, the First Amendment obviously excludes all 'governmental regulation of religious beliefs as such.'" *Id.* at 485 (quoting *Smith*, 494 U.S. at 877) (citation omitted).

44. 406 U.S. 205 (1972).

Supreme Court struck down a Wisconsin law forcing children of the Amish faith to attend public high school because the law lacked a compelling interest.[45] The *Smith* decision was criticized for limiting the *Sherbert-Yoder* test because it ruled that the compelling interest test did not apply to laws of neutral and general applicability.[46]

In 1993, Congress reacted to *Smith* by passing the Religious Freedom Restoration Act ("RFRA"),[47] which established a strict scrutiny test for any law that substantially burdened religious exercise.[48] This legislation reinstated the compelling interest test in *Sherbert* and *Yoder*.[49] As a result, RFRA undercut the Supreme Court's decision in *Smith*, which employed only a rational basis test for generally applicable laws.[50] Four years later, the Supreme Court weighed in on the constitutionality of RFRA in *City of Boerne v. Flores*.[51]

In *Boerne*, the Court struck down RFRA as applied to the states.[52] The Court held that under the Enforcement Clause of the Fourteenth Amendment,[53] legislation that attempts to alter the meaning of the Free Exercise Clause cannot be said to be enforcing the Free Exercise Clause.[54] The Court reasoned that when exercising its remedial powers, Congress' response to a violation must be both proportional and congruent.[55] RFRA was neither proportional nor congruent, and the Court held that it was a substantive change to the Fourteenth Amendment.[56]

---

45. *Id.* at 234.

46. *See* Note, *Religious Land Use in the Federal Courts under RLUIPA*, 120 HARV. L. REV. 2178, 2180 (2007) ("The scholarly and popular reaction to Justice Scalia's limiting of *Sherbert* was strongly negative.") (citing James E. Ryan, Note, Smith *and the Religious Freedom Restoration Act: An Iconoclastic Assessment*, 78 VA. L. REV. 1407, 1409–10, 1417 (1992)).

47. Religious Freedom Restoration Act of 1993, Pub. L. 103-141, 107 Stat. 1488 (codified as amended at 42 U.S.C. § 2000bb (2006)).

48. *See* 42 U.S.C. § 2000bb(a)(3). *See generally* IDES & MAY, *supra* note 43 ("For example, under RFRA, a neutral law banning the use of peyote could not be applied against a member of the Native American Church in the absence of a compelling state interest and a showing that the ban represented the least restrictive means to advance that interest.").

49. 42 U.S.C. § 2000bb(b)(1). In addition, the purposes of the chapter are "to guarantee . . . [the test's] application in all cases where free exercise of religion is substantially burdened; and to provide a claim or defense to persons whose religious exercise is substantially burdened by government." *Id.*

50. Emp't Div., Dep't of Human Res. of Or. v. Smith, 494 U.S. 872, 884–85 (1990).

51. 521 U.S. 507 (1997).

52. *Id.* at 532; *see* Gonzales v. O Centro Espirita Beneficente Uniao Do Vegetal, 546 U.S. 418, 423 (2006) (allowing a plaintiff to use RFRA against the federal government).

53. U.S. CONST. amend. XIV, § 5 ("The Congress shall have power to enforce, by appropriate legislation, the provisions of this article.").

54. *Boerne*, 521 U.S. at 519.

55. *Id.* at 519–20.

56. *Id.* at 532.

The Court pointed out some concerning deficiencies in RFRA. For example, a citizen could challenge any law under RFRA by alleging that the local government created a substantial burden on the free exercise of his or her religion.[57] Moreover, states were burdened by the requirement that their laws meet the strict scrutiny test.[58] The Court noted that RFRA lacked considerations of proportionality or congruence between the means adopted and the legitimate end to be achieved.[59] After the Court struck down RFRA, Congress limited its focus and passed RLUIPA in 2000.[60]

B. RLUIPA: Legislative Findings

The legislative findings leading up to RLUIPA's passage were the result of three years of hearings mainly focused on circumstances of religious discrimination in land use regulations.[61] The legislative history of RLUIPA helps shed light on the need for the enactment of the statute while also establishing guidelines for courts to interpret RLUIPA's vague statutory language. Part of the history of RLUIPA's language can be traced to the hearings on the Religious Liberty Protection Act ("RLPA")[62] in 1998

---

57. *Id.*

58. *Id.* at 534.

59. *Id.* at 533.

60. Religious Land Use and Institutionalized Persons Act of 2000, Pub. L. No. 106-274, 114 Stat. 803 (codified at 42 U.S.C. § 2000cc (2006)). RLUIPA and RFRA have notable differences. While RFRA essentially made almost every law affecting religious exercise jump through the hurdle of satisfying strict scrutiny, RLUIPA only presents this hurdle within the land use and institutionalized persons context. *See* Congregation Kol Ami v. Abington Twp., No. Civ.A. 01-1919, 2004 WL 1837037, at *10 (E.D. Pa. Aug. 17, 2004) (citation omitted) ("RLUIPA does not have '[s]weeping coverage.' . . . RLUIPA applies only to a very limited subject matter," that being "land use and regulations affecting institutionalized persons."). RFRA tried to make nearly every law justify its burden on religion by satisfaction of strict scrutiny. *Id.*; *see also* Cottonwood Christian Ctr. v. Cypress Redev. Agency, 218 F. Supp. 2d 1203, 1220 (C.D. Cal. 2002) ("RLUIPA only covers state action aimed at land use decisions and persons in jails or mental facilities.") (citing 42 U.S.C. §§ 2000cc–2000cc-1). The definition of religious exercise under RLUIPA, "any exercise of religion, whether or not compelled by, or central to, a system of religious belief," is broader than it is in RFRA. *RMCC I*, 612 F. Supp. 2d 1163, 1171 (D. Colo. 2009) (quoting 42 U.S.C. § 2000cc-5(7)(A)) (internal quotation marks omitted). The U.S. Supreme Court has typically limited religious exercise claims under RFRA to beliefs that are "central or fundamental to a person's religion." *Id.* at 1172 (citing Church of Lukumi Babalu Aye, Inc. v. City of Hialeah, 508 U.S. 520, 565 (1993)); *see* Grace United Methodist Church v. City of Cheyenne, 451 F.3d 643, 662 (10th Cir. 2006) (citing Thiry v. Carlson, 78 F.3d 1491, 1495 (10th Cir. 1996)). Further, as opposed to RFRA, RLUIPA's legislative findings have a more inherent focus on land use. *See Kol Ami*, 2004 WL 1837037, at *10.

61. RLUIPA was a bill based on hearings before the Senate Committee on the Judiciary and the House Subcommittee on the Constitution. 146 CONG. REC. 16,698 (2000) (joint statement of Sens. Hatch and Kennedy).

62. H.R. 1691, 106th Cong. (1999) (as passed by the House, July 15, 1999, received

and 1999.[63] The hearings evidenced the burdens that churches faced in building development and land use codes.[64]

The RLPA House Report cited testimony from the Subcommittee on the Constitution in the 105th and 106th Congresses and demonstrated the hardships churches faced when attempting to locate in residential zones.[65] Specifically, the testimony showed the hardship a church would face in searching to build a new structure in a residential zone because of that church's inability to convene in one household for religious practice.[66] Summarizing the hearing testimony, the House Report stated: "[T]he only way to build a church in a residential zone [was] to find several adjacent lots that [were] on the market simultaneously, buy them, and tear down the houses—an unfeasible strategy on its face."[67] Accordingly, churches were forced to locate in commercial districts.[68] However, the testimony explained that existing land use schemes only allowed for churches in residential zones.[69] Thus, although giving the appearance of being generous, these land use schemes were just the opposite.[70]

The report also cited zoning experts' opinions and history reflecting the separation of non-religious and religious use in the land use context.[71] For example, non-religious users such as commercial businesses have long been permitted in zones as of right, whereas churches have been regularly required to obtain a special use permit in those same zones.[72] In other instances, new churches were purposefully excluded from an entire city.[73] Such problems help illustrate land use laws' differential treatment of religious and non-religious uses.

The same arguments for land use legislation during the debate over RLPA are evident in the congressional record prior to the enactment of RLUIPA.[74] Noting how the First Amendment can be implicated in the land use context, the joint statement from Senators Orrin Hatch (R-UT) and Edward Kennedy (D-MA) provides that the right to build, rent, or buy physical space is necessary and at the core of the First Amendment right of assembly, a right which is frequently violated.[75] The record went on to

---

by the Senate, July 16, 1999, but never became law).

63. H.R. REP. NO. 106-219 (1999).

64. *Id.* at 18–19.

65. *Id.*

66. *Id.* at 18.

67. *Id.* (footnote omitted).

68. *Id.*

69. *Id.*

70. *Id.* at 18–19 (footnote omitted).

71. *Id.* at 18–24.

72. *Id.* at 19–20.

73. *Id.* at 19.

74. 146 CONG. REC. 16,698, 16,698–705 (2000) (joint statement of Sens. Hatch and Kennedy).

75. *Id.* at 16,698.

document a trend of zoning codes excluding churches, which were often the subjects of "covert" discrimination.[76] The difficulty of legislating in only those jurisdictions that engage in discrimination further increased the need for federal legislation.[77] Indeed, the congressional hearings revealed many constitutional grievances, rendering the situation ripe for RLUIPA's passage.[78]

II. 42 U.S.C. § 2000CC(B)(3)(B)—THE UNREASONABLE LIMITATIONS PROVISION

RLUIPA's unreasonable limitations provision is found in the discrimination and exclusion subsection of the statute.[79] The discrimination and exclusion subsection in its entirety reads as follows:

> (b) Discrimination and exclusion
>
> (1) Equal terms
>
> No government shall impose or implement a land use regulation in a manner that treats a religious assembly or institution on less than equal terms with a nonreligious assembly or institution.
>
> (2) Nondiscrimination
>
> No government shall impose or implement a land use regulation that discriminates against any assembly or institution on the basis of religion or religious denomination.
>
> (3) Exclusions and limits
>
> No government shall impose or implement a land use regulation that—
>
> (A) totally excludes religious assemblies from a jurisdiction; or
>
> (B) unreasonably limits religious assemblies, institutions, or structures within a jurisdiction.[80]

Section 2000cc(b)(3)(B), the unreasonable limitations provision, is specifically encompassed in the exclusion and limits section.[81] Generally, this particular section has addressed exclusionary zoning codes and restrictive land use practices.[82] An example of exclusion per § 2000cc(b)(3)(A) is where a zoning code disallows religious institutions but allows other organizations. The unreasonable limitations provision seems

---

76. *Id.* at 16,697–98. *But see* Patricia E. Salkin & Amy Lavine, *The Genesis of RLUIPA and Federalism: Evaluating the Creation of a Federal Statutory Right and Its Impact on Local Government*, 40 URB. LAW. 195, 256–57 (2008) (noting that critics have challenged the legitimacy of the data Congress relied on to demonstrate that "widespread" discrimination against religious institutions was taking place).

77. 146 CONG. REC. 16,698, 16,699 (2000).

78. *See id.* at 16,698–703.

79. 42 U.S.C. § 2000cc(b)(3) (2006).

80. *Id.* § 2000cc(b).

81. *Id.*

82. *See infra* Part II.B–C.

simple by its plain language. However, questions about how far and to what degree a government can limit religious institutions remain unresolved and are determined by courts on a case-by-case basis.

A. Legislative History of Unreasonable Limits

The legislative history of RLUIPA is important to understand this provision because it provides insight on what factors should be used to determine unreasonable limits.[83] The joint statement of Senators Hatch and Kennedy from the congressional record explains that the exclusion or unreasonable limitations provision "enforces the Free Speech Clause as interpreted in *Schad v. Borough of Mount Ephraim*, which held that a municipality cannot entirely exclude a category of [F]irst [A]mendment activity."[84] Further, the record explains that this exclusion and limits section protects "the right to assemble for worship or other religious exercise under the Free Exercise Clause."[85]

With respect to unreasonable limitations, the legislative history of 42 U.S.C. § 2000cc provides that "[w]hat is reasonable must be determined in light of all the facts, including the actual availability of land and the economics of religious organizations."[86] Similarly, the House Report from RLPA explains that reasonableness under § 3(b)(1)(D) of RLPA, which corresponds to the unreasonable limitations provision of RLUIPA, "must be determined in light of all facts, including the physical and financial availability of land to religious organizations."[87]

These statements, albeit rather short, provide a general framework that courts can use to ascertain congressional intent. For example, a federal district court in Florida in *Chabad of Nova, Inc. v. City of Cooper City*, in determining whether a law violated the unreasonable limitations provision, looked to the legislative history in assessing whether a religious assembly had reasonable alternatives to build elsewhere.[88] In *Chabad of Nova*, the plaintiff, a Jewish orthodox outreach center, sued the city defendant using RLUIPA's unreasonable limitations provision because of the plaintiff's inability to rent space and purchase property within the city.[89] The religious claimant in *Chabad of Nova* fashioned an argument based on the analysis

---

83. 146 CONG. REC. 16,698–703.

84. *Id.* at 16,700 (citing Schad v. Borough of Mount Ephraim, 452 U.S. 61 (1981)).

85. *Id.*

86. 146 CONG. REC. E1,563 (daily ed. Sept. 22, 2000) (statement of Rep. Canady on the Religious Land Use and Institutionalized Persons Act of 2000).

87. H.R. REP. NO. 106-219, at 29 (1999) (referencing RLPA, H.R. 1691, 106th Cong. (1999) (passed by the House, July 15, 1999, received by the Senate, July 16, 1999, but never became law).

88. Chabad of Nova, Inc. v. City of Cooper City, 575 F. Supp. 2d 1280, 1289 (S.D. Fla. 2008).

89. *Id.* at 1285.

proposed by Congress in RLUIPA's legislative history as to how courts should determine reasonableness.[90] The claimant asked the court to evaluate whether the zoning scheme at issue provided enough sites for a reasonable opportunity for religious expression.[91]

The specific land use ordinance in *Chabad of Nova* prohibited religious assemblies from locating in commercial areas.[92] With respect to the purchase of property, the plaintiff presented evidence that it would need to buy an average of five properties just to meet the defendant's frontage requirements in a residential zone.[93] In 2006, the additional cost that would have been incurred to meet the frontage requirements ranged anywhere from $880,000 to over $2.5 million, rendering religious assemblies unequal market participants as compared to the ordinary land user.[94] The court deemed Cooper City's ordinance a violation of the unreasonable limitations provision because of the limited availability of property for religious assemblies to locate in the city, the inflated costs to locate, and the more stringent requirements imposed on religious assemblies versus other similar uses.[95]

Chabad of Nova is important because it shows how arguments based on reasonable alternatives or locations for religious expression factor into the unreasonable limitations provision analysis. If a religious institution is treated unreasonably in its effort to locate in a particular zone, and there are not adequate alternatives in that particular area, then that institution is likely unreasonably limited under RLUIPA.

Chabad of Nova also demonstrates that RLUIPA's legislative history is important for courts when deciding cases. However, the problem courts face is that the legislative history on the unreasonable limitations provision provides only a broad statement as to how to assess reasonableness. The end result is that courts are left with the ultimate burden to make sense of RLUIPA's statutory language. A court must determine just how much religious freedom a government can limit in the land use context. Different circuits have adopted conflicting interpretations of RLUIPA's substantial burden provision.[96] The difficulty of interpreting the unreasonable limitations provision will likely persist as more RLUIPA suits are brought.

---

90. *Id.* at 1289–90. This may have been a reason why the defendant did not argue against this point. *See id.* at 1289. The court noted that such an analysis was consistent with the legislative history of RLUIPA. *Id.*

91. *Id.*

92. *Id.* at 1283.

93. *Id.* at 1290.

94. *Id.*

95. *Id.*

96. *See, e.g*, MacLeod, *supra* note 6, at 53–62 (noting a Seventh Circuit test, a Ninth Circuit test, and the standard used by courts in other circuits for construing the substantial burden provision).

B. Case Law: The Unreasonable Limitations Provision

The unreasonable limitations provision states that "[n]o government shall impose or implement a land use regulation that . . . unreasonably limits religious assemblies, institutions, or structures within a jurisdiction."[97] A Pennsylvania federal district court, following Supreme Court precedent, has weighed in on the constitutionality of the unreasonable limitations provision.[98] In *Freedom Baptist Church v. Middletown*, the court stated that RLUIPA's unreasonable limitations provision "codifies existing Supreme Court Equal Protection jurisprudence under the Fourteenth Amendment" and does not violate the First Amendment.[99] The court used the Supreme Court decision in *City of Cleburne v. Cleburne Living Center, Inc.* to justify this holding.[100]

In *Cleburne*, a land use regulation required the operators of a home for the mentally handicapped to obtain a special use permit in an area that allowed a number of other uses as of right.[101] The Court concluded that this was "irrational prejudice against the mentally retarded" and affirmed the lower courts' decisions to strike down the ordinance on Equal Protection grounds.[102] The *Middletown* court discerned from *Cleburne* that Congress "did no more than codify settled Supreme Court standards" with respect to the unreasonable limitations provision.[103]

Aside from the issue of constitutionality, courts such as the Middle District of Florida in *Konikov v. Orange County* have addressed the unreasonable limitations provision by assessing the level of protection religious institutions should receive in the land use context.[104] In *Konikov*, the court reasoned that religious institutions are not entitled to immunity from land use regulations[105] by stating that "[the exclusion and limits] provision suggests that Congress contemplated that religious assemblies could be reasonably limited within a jurisdiction."[106] Quoting from RLUIPA's legislative history, the court stated that RLUIPA does not "relieve religious institutions from applying for variances, special permits or exceptions, hardship approval, or other relief provisions in land use

---

97. 42 U.S.C. § 2000cc(b)(3) (2006).

98. Freedom Baptist Church of Delaware Cnty. v. Twp. of Middletown, 204 F. Supp. 2d 857, 870–71 (E.D. Pa. 2002).

99. *Id.* at 871.

100. *Id.* (citing City of Cleburne v. Cleburne Living Ctr., 473 U.S. 432, 447–48 (1985)).

101. *Cleburne*, 473 U.S. at 435, 450.

102. *Id.* at 450.

103. *Middletown*, 204 F. Supp. 2d at 871.

104. Konikov v. Orange Cnty., Florida, 302 F. Supp. 2d 1328, 1345–46 (M.D. Fla. 2004), *aff'd in part, rev'd in part*, 410 F.3d 1317 (11th Cir. 2005).

105. *Id.* at 1346.

106. *Id.*; *see also* Salkin & Lavine, *supra* note 76, at 251 (citing *Konikov* when discussing reasonable limitations with respect to the provision).

regulations, where available without discrimination or unfair delay."[107] This analysis suggests that a special use permit or other similar requirement is a reasonable limitation within the context of the unreasonable limitations provision.[108]

The Seventh Circuit agreed with this reasoning in *Vision Church, United Methodist v. Village of Long Grove*.[109] The appellant Vision Church tried to use the unreasonable limitations provision in its favor by arguing that it was unreasonably limited due to the Village Board having unbridled discretion in issuing special use permits.[110] The court disagreed, stating that "this is not a case where the 'state [has] delegate[d] essentially standardless discretion to nonprofessionals operating without procedural safeguards.'"[111] The Seventh Circuit reasoned that even if the zoning regulations granted the board undue discretion, this discretion would not constitute a violation of unreasonable limitations.[112] The court held that the special use permit requirement for a church did not amount to an unreasonable limitation and linked such permits to legitimate municipal goals.[113] Thus, the court concluded, the requirement that a church obtain a special use permit, variance, or other exception does not violate the unreasonable limitations provision unless similar requirements are not imposed on secular institutions.[114]

C. RMCC II *and its Importance for RLUIPA*

In 2010, the Tenth Circuit decided *Rocky Mountain Christian Church v. Board of County Commissioners of Boulder County*.[115] The case is important for a number of reasons.[116] First, the case is a more recent

---

107. *Konikov*, 302 F. Supp. 2d at 1345–46 (citing 146 CONG. REC. S7,774, 7,776 (daily ed. July 27, 2000) (joint statement of Sens. Hatch and Kennedy)).

108. *See id.*

109. 468 F.3d 975, 990–91 (7th Cir. 2006).

110. *Id.* at 990.

111. *Id.* (quoting Saints Constantine & Helen Greek Orthodox Church, Inc. v. City of New Berlin, 396 F.3d 895, 900 (7th Cir. 2005)) (alteration in original).

112. *Id.* at 991. *But see* Salkin & Lavine, *supra* note 76, at 251 ("[R]eligious restrictions may be unreasonable if they operate according to the unbridled discretion of local officials, but in such a case a plaintiff must do more than raise a facial challenge—it must show that such discretion, as used, was in fact unreasonable.") (referencing *Vision Church*, 468 F.3d at 990–91).

113. *Vision Church*, 468 F.3d at 991.

114. *Id.*

115. *RMCC II*, 613 F.3d 1229, 1230 (10th Cir. 2010).

116. *See generally* Clay Evans, *High Stakes Land Fight: Boulder County Right to Fight Rocky Mountain Christian Church Expansion*, BOULDER DAILY CAMERA (July 20, 2010, 11:41 PM), http://www.dailycamera.com/ci_15561981 ("The case . . . is about the constitutionality of RLUIPA—whether Congress can expand religious protections beyond the First Amendment to essentially give religious entities more rights than other entities in

RLUIPA case and involves both categories of RLUIPA's land use portion of the statute: the substantial burden and the discrimination and exclusion sections.[117] Second, the case helps highlight the vague contours of RLUIPA's provisions, particularly with respect to their application in an expansion context.[118] Finally, the case is important because of the parties involved. A church, emerging as the largest in its county, desired to continue its expansion but faced opposition from a county known for its reputable land use goals.[119] The direction the Colorado district court and Tenth Circuit took in *RMCC I* and *II* helps not only to assess the current state of RLUIPA's provisions but also offers an opportunity to analyze future implications, an analysis that will follow later in this Comment. The case's facts are important to gain an understanding of this upcoming analysis.

In 1978, Boulder County implemented a comprehensive plan[120] to curb urban sprawl and preserve the county's rural character by maintaining open space.[121] Boulder County changed its land use code in 1983 to further implement this comprehensive plan.[122] In 1994, the County amended its zoning code to require all facilities with 100 people or more in an Agricultural District to apply for a special use permit.[123] Rocky Mountain Christian Church ("RMCC"), founded in 1984, was located in such a district.[124] RMCC was permitted to expand without any special use applications until the 1994 amendments to the zoning code.[125]

---

challenging local zoning laws. . . . This is precisely the kind of case that should wind up before the high court.").

117. *RMCC II*, 613 F.3d at 1240.

118. *Id.* at 1234.

119. *Id.* at 1233–34.

120. *See also supra* note 26 (detailing the purpose of comprehensive plans). *See generally* Daniel J. Curtin, Jr., Ramapo*'s Impact on the Comprehensive Plan*, 35 URB. LAW. 135, 147 (2003) ("The General [i.e. Comprehensive] Plan represents the most comprehensive statement of the community's welfare relative to environmental and land use matters.").

121. *RMCC II*, 613 F.3d at 1233.

122. *Id.*

123. *Id.* Boulder's Land Use Code divides the county into thirteen zoning districts, which includes Agricultural Districts. *Id.* "The special use process entails an application to the County's land use office, and public hearings and votes before both the Planning Commission and the County Commission." *Id.* For an exposition of the role of special use permits, see *Zylka v. City of Crystal*, 167 N.W.2d 45, 48 (Minn. 1969) ("[Special use permits] are designed to meet the problem which areas where certain uses, although generally compatible with the basic use classification of a particular zone, should not be permitted to be located as a matter of right in every area included within the zone because of hazards inherent in the use itself or special problems which its proposed location may present.").

124. *RMCC II*, 613 F.3d at 1233–34.

125. *Id.* at 1234. Thereafter, RMCC was grandfathered into the land use code. *Id.*

In 1997, RMCC sought to make its structure a conforming use and build a 54,000 square foot school by applying for a special use permit.[126] The County Commission consented to this permit despite disagreement with the County's land use staff.[127] In 2002, the County, again over the opposition of the land use staff, approved another special use application for a 120-student expansion and placement of temporary modular units on RMCC's campus.[128] The 2004 special use application in contention sought a 28,000 square foot gymnasium, a 6,500 square foot chapel, gallery space to connect the buildings, and an increase in its main worship building seating by 150 seats.[129]

The County's land use staff found that the 2004 application conflicted with goals of its comprehensive plan because the expansion was "incompatible with the surrounding area, an over-intensive use of the land, likely to cause undue traffic congestion, and likely detrimental to the welfare of the residents of Boulder County."[130] The land use staff used an unconventional calculation to find an over-intensive use of land.[131] The typical method was that "a proposed use [was] not over-intensive if less than 50% of the site's surface area would be covered by a structure or a parking lot."[132] RMCC's expansion resulted in 35% coverage of their land.[133] Despite RMCC being under the typical threshold, the land use staff found an over-intensive use of land based on the fact that the expansion would double the church's square footage and significantly increase its parking area.[134] In light of this evidence and other factors, the Planning Commission voted against the application, and the County Commission rendered a final partial denial.[135]

Thereafter, RMCC brought a RLUIPA claim, and the case proceeded to trial.[136] The jury returned verdicts in favor of the plaintiff based on the

---

126. *Id.*

127. *Id.*

128. *Id.*

129. *Id.* This plan was based on a projected expansion over a period of twenty years. *Id.* The initial application originally sought approval for 150,200 additional square feet. *Id.* According to Boulder County Attorney's Office, the county had approved five previous requests from RMCC to expand during the time period of 1993-2003. Evans, *supra* note 116. Since the adoption of this special-use requirement, every one of forty-five applications by churches had been approved except for the 2004 special use application in contention. *Id.*

130. *RMCC II*, 613 F.3d at 1234.

131. *Id.* at 1234–35.

132. *Id.*

133. *Id.* at 1235.

134. *Id.*

135. *Id.* The application was denied except for the increased seating capacity of 150 seats and the 10,000 square foot building replacement for the modular units. *Id.* At the public hearing where this vote took place, an RMCC consultant was greeted by a commissioner who said, "Rosi, you can bring in your Christians now." *Id.*

136. *Id.* During this RLUIPA challenge, legislators from Boulder, Colorado, issued a

equal terms, substantial burden, and unreasonable limitations provisions of RLUIPA.[137] In its review of the jury finding a violation of unreasonable limitations, the district court noted that the literal language of the unreasonable limitations provision is in plural.[138] The court reasoned that the provision requires more than one religious organization to be unreasonably limited because the provision mentions "religious assemblies, institutions, or structures."[139] Thus, to violate RLUIPA's unreasonable limitations provision, the religious exercise of both RMCC and other religious institutions or assemblies needed to be deprived.[140]

The district court reviewed evidence on the unreasonable limitations finding[141] based largely on RMCC's expert witness, who had previously worked for Boulder County's land use department for twelve years.[142] The witness testified that "the county's regulations had become more difficult over the years, and that any church of any size must go through either a site plan review or a special use review."[143] The same witness testified that the County often placed conditions on churches' special use applications that "reduce[d] either the number of people permitted or the number of square feet permitted in a facility."[144] Based on these findings, the district court found sufficient evidence that churches other than RMCC were unreasonably limited by Boulder County's land use regulations.[145] With respect to the equal terms and substantial burden claims, the district court found there to be sufficient evidence for RLUIPA violations of both provisions.[146]

RMCC appealed, and the Tenth Circuit reviewed the factual record to determine whether there was sufficient evidence for the jury to make its findings based on the unreasonable limitations, equal terms, and substantial burden provisions.[147] As to unreasonable limitations, the court considered

---

resolution to their congressional representative requesting an invalidation of RLUIPA. Salkin & Lavine, *supra* note 76, at 253 (footnote omitted).

137. *RMCC II*, 613 F.3d at 1235.

138. *RMCC I*, 612 F. Supp. 2d 1163, 1176 (D. Colo. 2009).

139. *Id.* at 1176–77.

140. *Id.* at 1176.

141. *Id.* at 1176–77.

142. *Id.* The witness worked as a consultant to churches in the area regarding the process of special use permits. *Id.*

143. *Id.* at 1177. However, she did agree that "anyone in Boulder County who applies for permission to do something on their property faces the same regulations and limitations." *Id.*

144. *Id.*

145. *Id.*

146. *Id.*

147. *RMCC II*, 613 F.3d 1229, 1235 (10th Cir. 2010). On June 17, 2010, Boulder County asked the federal appeals court to reconsider its upholding of the jury's determination on May 17, 2010. That decision was affirmed in a seven-page opinion on July 19, 2010. *See id.* at 1230. The Tenth Circuit did not address the substantial burden provision

evidence that a witness approached the County Commissioner with the prospect of building a synagogue and received a response saying that "there will never be another mega church . . . in Boulder County."[148] Further, RMCC's expert explained at trial that the County's land use plan has made it "more difficult for churches to operate in Boulder County."[149] In RMCC's case, the land use staff engaged in the uncharacteristic practice of changing both lot sizes and building square footages on RMCC's report after the Planning Commission meeting.[150] The Tenth Circuit found this evidence "more than adequate" for a jury to find a violation of unreasonable limitations.[151]

The Tenth Circuit also found sufficient evidence at trial for a violation of RLUIPA's equal terms provision, § 2000cc(b)(1), "impos[ing] or implement[ing] a land use regulation in a manner that treats a religious assembly or institution on less than equal terms with a nonreligious assembly or institution."[152] Because the County approved a project similar to RMCC's plan, RMCC argued that the County's denial of its plan was a violation of the equal terms provision.[153] Upon reviewing all of the evidence, the Tenth Circuit affirmed the district court's holding in favor of RMCC on this issue.[154]

---

since a violation of the other provisions was enough to issue the district court's injunction. *Id.* at 1239.

148. *Id.* at 1238.

149. *Id.*

150. *Id.*

151. *Id.* at 1239. Further evidence was introduced by RMCC to resolve the County's concerns with its proposed development. *Id.* at 1238. For instance, to minimize its visual impact on the neighborhood around it, RMCC hired a surveyor "to confirm that the church's proposed landscaping and berming would block all views of the expanded building from every neighboring home." *Id.*

152. *Id.* at 1236–38. There is some debate about the proper standard of review for RLUIPA's equal terms provision. *See* Salkin & Lavine, *supra* note 76, at 247 ("The more vexing dilemma is how to determine whether a religious use has been treated on less than equal terms than a secular use.").

153. *RMCC II*, 613 F.3d at 1236–37.

154. *Id.* at 1238. The Tenth Circuit did not address the substantial burden provision since a violation of the other provisions was enough to issue the district court's injunction. *Id.* at 1239.

III. RLUIPA'S UNREASONABLE LIMITATIONS PROVISION AND CASES ANALYZED

A. An Analysis of RMCC II's *Outcome with Regard to the Unreasonable Limitations Provision*

The district court and the Tenth Circuit both found that Boulder County had violated the unreasonable limitations provision in *RMCC I* and *II*.[155] Boulder County's argument for the maintenance of land use goals might have been taken more seriously absent the alleged unfairness that took place. Regardless, the case presents the interesting question of whether RMCC's 2004 application was an instance of a local government going out of its way to place more than necessary restrictions on a religious claimant or whether that government was merely trying to adhere to the goals of its comprehensive plan.

On one hand, Boulder County could point to the fact that it had accommodated religious institutions in the past[156] as mitigating evidence suggesting that the County did not unreasonably limit RMCC's proposed land use in this case. The County had granted a number of RMCC's previous special use applications.[157] Also, an expert witness previously employed by the County agreed that she had not been aware of the denial of any church's special use application except the one at issue.[158] Thus, Boulder County may have indeed been trying to adhere to its comprehensive plan goals when it denied RMCC's 2004 application.

However, the problem lies in the way that the County treated RMCC's 2004 application. In addition to the "mega church" comment by the County Commissioner,[159] the County's processing and determining of the church's application showed that RMCC was potentially unreasonably limited.[160] Also, the testimony at trial revealed that Boulder County churches faced a pattern of difficulty and limitations in the special use application process.[161] This evidence permitted a finding that the County created difficulties for multiple religious organizations in the land use assessment process, thereby demonstrating that one or more religious institutions were unreasonably limited in terms of the provision.

The Tenth Circuit's analysis reiterates this conclusion. In *RMCC II*, the Tenth Circuit cited testimony that churches were having a difficult time operating in Boulder County.[162] This testimony, coupled with other

155. *See supra* Part II.C.
156. *See RMCC II*, 613 F.3d at 1234.
157. *Id.*
158. *RMCC I*, 612 F. Supp. 2d at 1177.
159. *RMCC II*, 613 F.3d at 1238.
160. *Id.* at 1234–35.
161. *Id.* at 1238.
162. *Id.*

evidence and the adverse treatment to RMCC, helped the court conclude that there was "more than adequate" evidence for a finding of unreasonable limits.[163] A lesson that other municipalities can take away from *RMCC II* is to treat every religious institution fairly, particularly on each of their permit applications, to avoid a RLUIPA unreasonable limitations claim or a general RLUIPA claim. Further, municipalities should use the same standards for religious institutions as done for other secular institutions to avoid a RLUIPA challenge.[164]

B. Boulder County's Land Use Policy and Rationale for Excluding Churches

Some land use experts are rightfully puzzled about the recent *RMCC II* decision.[165] As Patricia Salkin, a land-use law expert at Albany Law School in New York, said, "Boulder has a very good reputation for being a community that is solid when it comes to comprehensive planning land-use regulations. They are often held up as a model community."[166]

Boulder has a history of restricting growth to preserve its character and high quality of life.[167] While Colorado experienced a 30.6% population growth in the 1990s, Boulder's population increased by only 13.6%.[168] Part of the reasoning behind such limited growth was to combat the negative effects of urban sprawl.[169] In 1998, Boulder adopted a Residential Growth Management System ("RGMS").[170] This system was adopted to address unmanaged growth in the city and was aimed at preventing "more crime, pollution, urban decay, and stress on the infrastructure and public facilities."[171] The RGMS aspires to create "a rate of growth in the city that will assure the preservation of its unique environment and its high quality of life."[172] More broadly, the title section of Boulder's land use regulations in which the RGMS is written aims to "[p]reserve the character and stability of neighborhoods and conserve property values by encouraging the most

---

163. *Id.* at 1238–39.

164. *See Vision Church*, 468 F.3d 975, 991 (7th Cir. 2006).

165. *See* Laura Snider, *Boulder County's Denied Appeal in Church Expansion Garners National Attention*, BOULDER DAILY CAMERA (May 21, 2010, 8:03 PM), http://www.dailycamera.com/longmont-news/ci_15138209.

166. *Id. See generally* Katharine J. Jackson, *The Need for Regional Management of Growth: Boulder, Colorado, as a Case Study*, 37 URB. LAW. 299 (2005) (providing statistics and literature that support Boulder's reputable land use goals).

167. *See* Jackson, *supra* note 166, at 299 ("Although it is a popular place to live, in the 1990s Boulder's population grew at a rate much slower than that of Colorado as a whole.").

168. *Id.*

169. *Id.*

170. *Id.* at 312.

171. *Id.*

172. BOULDER, COLO., REV. CODE, § 9-14-1(b) (2007), *available at* http://www.colocode.com/boulder2/chapter9-14.htm.

appropriate uses of land within zoning districts."[173] Thus, both statistics and Boulder's land use policy demonstrate the city's dedication to maintaining an ideal growth rate.

In *RMCC II*, Boulder County's dedication to an ideal growth rate was reflected in its comprehensive plan that aimed to maintain open space and curb sprawl.[174] Rationalizations for RMCC's adverse treatment are tied to the implementation of this plan and are founded upon the overall effects of a church's expansion.[175] At certain times of the day, churches create big bursts of traffic that can lead to noise and congestion. This noise and congestion can adversely affect the quality of life for community residents.[176] Further, the city bears the responsibility for providing solutions to increased traffic problems, which might not be feasible in some instances due to city infrastructure. The County cited curbing these types of effects as a reason for the denial of RMCC's application.[177] The County, therefore, might have been justified in denying RMCC's 2004 application.

However, local governments sometimes oppose land use applications for religious institutions because they generally do not pay property taxes.[178] An expanding church that continues to take up more land forecloses the opportunity for a city to bring in additional revenue if other potential users of that land would pay taxes. Boulder County's opposition to RMCC could have been due, in part, to the prospect of losing additional tax revenue. But, in light of the history and reputation of Boulder's land use policy and goals, this rationale is unlikely the reason why RMCC's 2004 application faced hostility from Boulder County. Moreover, since Boulder County's enactment of its special-use requirement for large buildings in 1994, all forty-five applications by churches had been approved prior to RMCC's 2004 special use application.[179] Thus, this evidence further suggests that Boulder County was trying to adhere to its comprehensive plan rather than express hostility against RMCC.

---

173. *Id.* § 9-1-1(e), *available at* http://www.colocode.com/boulder2/chapter9-1.htm.

174. *See RMCC II*, 613 F.3d 1229, 1233 (10th Cir. 2010); Jackson, *supra* note 166, at 299.

175. *See RMCC II*, 613 F.3d at 1233–34; Jackson, *supra* note 166, at 299.

176. *See generally* Weinstein, *supra* note 31, at 4 ("As church activities expand to twelve or more hours per day seven days a week, neighbors become increasingly concerned about the negative effects of the increased traffic, parking, noise, and late-night activity on property values.").

177. *RMCC II*, 613 F.3d at 1234.

178. *See* Weinstein, *supra* note 31, at 5 ("Local officials may also be concerned [with respect to large houses of worship] about erosion of the city's tax base if too much property is acquired by tax-exempt religious institutions.").

179. Evans, *supra* note 116.

C. *Unreasonable Limitations:* RMCC II *in Light of Other Cases*

1. Unreasonable Limitations and Expansion

The cases discussed to this point provide a better perspective about RLUIPA's unreasonable limitations provision but still leave unresolved questions.[180] For example, the Colorado district court in *RMCC I* provided a unique analysis of the *literal* language of the provision by explaining that the language of the provision is plural, meaning that more than one religious institution must be unreasonably limited to violate the provision.[181] Further, the facts of *RMCC I* showed that instances of adverse treatment of one particular applicant coupled with a land use process hostile to religious institutions constitutes grounds for an unreasonable limitations claim.[182] However, certain issues, such as the relation of the unreasonable limitations provision to expansion, are still open to question. Recent cases not only illustrate this problem but also raise the nascent issue of the absence of limitations in RLUIPA.

In *Konikov*, a Florida federal district court reasoned that the unreasonable limitations provision "suggests that Congress contemplated that religious assemblies could be reasonably limited within a jurisdiction."[183] However, within the context of expansion, it is unclear just how far a religious institution can be limited.[184] In *RMCC II*, the church exhibited a pattern of continuous growth. This is evidenced by Boulder County's previous grants of RMCC's special use applications requesting escalations in building square footages and growth of its student body.[185] The County, citing an over-intensive use of land in 2004, then halted RMCC's expansion.[186] However, at trial, RMCC prevailed.[187] The County's stated reason for denying the permit, the maintenance of land use goals, was rejected, whether or not it was legitimate.[188]

From the district court and the Tenth Circuit's analyses in *RMCC I* and *II*, one can discern no reasoning that clarifies just how far a church's expansion can go with respect to the unreasonable limitations provision. Further, *Chabad of Nova* and *RMCC I* suggest that an unnecessarily restrictive assessment process or land use code must be in place for this

180. *See supra* Part II.B–C.

181. *RMCC I*, 612 F. Supp. 2d 1163, 1176 (D. Colo. 2009).

182. *See supra* Part II.C.

183. Konikov v. Orange Cnty., Florida, 302 F. Supp. 2d 1328, 1328 (M.D. Fla. 2004), *aff'd in part, rev'd in part*, 410 F.3d 1317 (11th Cir. 2005).

184. *See* MacLeod, *supra* note 6, at 50 ("[T]he mega-church appears to be the great source of contention among those who take an interest in RLUIPA.").

185. *RMCC II*, 613 F.3d 1229, 1234 (10th Cir. 2010).

186. *Id.* at 1234–35.

187. *RMCC I*, 612 F. Supp. 2d 1163, 1167 (D. Colo. 2009).

188. *RMCC II*, 613 F.3d at 1240.

provision to apply. Thus, a lingering question is whether the unreasonable limitations provision even applies if a religious institution is denied an expansion opportunity.

In 2010, a Pennsylvania district court in *West Pikeland* reasoned that the provision is not meant to examine restrictions placed on individual landowners.[189] In *West Pikeland*, the plaintiff asserted that a limitation on its building size was unreasonable.[190] Upon finding no violation of the unreasonable limitations provision, the district court focused on the literal language of the provision.[191] Specifically, the court focused on language mandating that religious institutions cannot be unreasonably limited "within a jurisdiction."[192] The court's focus on the phrase "within a jurisdiction" suggests that the provision is not meant to address particularized limitations on an individual plot of land but is intended to prevent "municipalities from broadly limiting where religious entities can locate."[193]

In *West Pikeland*, the plaintiff's land was also encumbered by a restrictive covenant, a fact which contributed to the court's decision to rule against the plaintiff on its unreasonable limitations claim.[194] The court reasoned that it is not an unreasonable limitation to require compliance with a restrictive covenant on land.[195] Nevertheless, although the restrictive

---

189. *West Pikeland*, No. 09-cv-1626, 2010 WL 2635979, at *17 (E.D. Pa. June 25, 2010). In *West Pikeland*, the plaintiff bought a plot of land in Chester Springs, Pennsylvania, located within the township of West Pikeland. *Id.* at *2. This plot of land was part of a much larger lot that was subdivided by the prior owners. *Id.* West Pikeland approved this subdivision and imposed limitations by restrictive covenants on any future development. *Id.*

After acquisition of the subject property, in September of 2001, the plaintiff filed a request for a special exception to build a 5,000 square foot structure to use "as a church or similar place of worship." *Id.* at *3. In April of 2002, the Board granted the plaintiff's special use application subject to certain conditions such as providing a parking plan and modifying the driveway of the property. *Id.* In 2008, the plaintiff filed a conditional-use application for the building of a 26,370 square-foot temple and a 9,100 square-foot auxiliary building. *Id.* Among its other findings, the Zoning Board found that the plaintiff's 2008 proposal included development outside of the cross-hatched markings on the subdivision plan, which was violative of the restrictive covenant. *Id.* However, the Zoning Board recognized the applicant's right to worship and approved the application on the condition that the plaintiff submit another plan that limited development to a single structure not exceeding 5,000 square feet within the cross-hatched areas of the subdivision plan. *Id.* Thereafter, the plaintiff filed, among other claims, a RLUIPA complaint. *Id.* at *4.

190. *Id.* at *17.

191. *Id.* at *17–18.

192. *Id.*

193. *Id.*; *see* 146 CONG. REC. 16,698, 16,700 (2000) (joint statement of Sens. Hatch and Kennedy) (explaining that the exclusion or unreasonable limitations provision "enforces the Free Speech Clause as interpreted in *Schad v. Borough of Mount Ephraim*, 425 U.S. 61 (1981), which held that a municipality cannot entirely exclude a category of First Amendment activity.").

194. *West Pikeland*, 2010 WL 2635979, at *17.

195. *Id.* In *West Pikeland*, the Township imposed restrictive covenants limiting future

covenant was part of the court's reasoning in ruling against the plaintiff's unreasonable limitations claim, the court's overall analysis is instructive for the applicability of the provision. The court suggested that the unreasonable limitations provision is limited to preventing municipalities from engaging in practices that limit where religious entities can locate.[196]

This basic premise behind the provision—that it is meant to prevent municipalities from broadly limiting where religious entities can locate—is correct, but it is too narrow as to what the provision might cover.[197] In light of *RMCC II*, the provision can be problematic in an expansion context. Boulder County placed limitations on RMCC, a church wanting to expand, by creating unnecessary restrictive limitations on RMCC during its 2004 application assessment, such as the alteration of its land use report.[198] *RMCC II* illustrates a situation where a church was not seeking to locate in a particular jurisdiction but still successfully proved an unreasonable limitations claim.[199] Thus, unreasonable limitations on religious institutions or assemblies can apply in various contexts.

Further, *RMCC II* shows that a court may consider a broad spectrum of evidence in assessing a violation of unreasonable limitations and need not confine its examination to how a land use authority has deterred a religious claimant's location effort. Although the Tenth Circuit considered that Boulder County's land use scheme had the effect of leaving few sites for church construction, it also considered the county's restrictive practices employed during RMCC's 2004 application.[200] Additionally, the court reviewed evidence that another congregation ran out of money during the County's special use application process, forcing it to abandon its building project.[201] This evidence, coupled with RMCC's "disparate" treatment during its 2004 application assessment, was "more than adequate" for a jury to find a violation of the unreasonable limitations section.[202]

2. Expansion Opportunities Compared Under the Substantial Burden Provision Versus the Unreasonable Limitations Provision

The issue of expansion and whether any type of land use limit can be imposed on a church arises more under RLUIPA's substantial burden provision than under the unreasonable limitations provision.[203] This is

---

development after the prior owners of plaintiff's lot subdivided their property. *Id.* at *2.

196. *See supra* note 193 and accompanying text.

197. This premise is consistent with RLUIPA's legislative history describing the purpose of the unreasonable limitations provision.

198. *RMCC II*, 613 F.3d 1229, 1238 (10th Cir. 2010).

199. *Id.* at 1238–39.

200. *Id.* at 1238.

201. *Id.*

202. *Id.* at 1239.

203. *See* Cottonwood Christian Ctr. v. Cypress Redev. Agency, 218 F. Supp. 2d 1203

because RLUIPA's exclusion and limits section encompasses the unreasonable limitations provision.[204] Case law demonstrates that this section applies to restrictive practices in the land use process more than an expansion effort. Also, the unreasonable limitations provision's plural requirement, requiring more than one incidence of unreasonable limits, is an impediment to religious claimants who experience only one hostile act.[205] In this instance, it might be easier for an aggrieved party to assert a RLUIPA substantial burden claim because the substantial burden provision does not require more than one act of unfair treatment.[206] Thus, a religious claimant only equipped with evidence of its own mistreatment will likely not look to the unreasonable limitations provision first.

Although claimants more readily use RLUIPA's substantial burden provision when denied an expansion opportunity, a land use authority should still be ready to face both a RLUIPA unreasonable limitations claim and a substantial burden claim. This is because religious claimants, such as the one in *West Pikeland*, are still under the impression that denial of an application for expansion can violate the unreasonable limitations provision.[207] Further, when pleaded properly, the provision can create problems for a land use authority when it denies a religious claimant's expansion. Specifically, a religious claimant can allege a violation of the provision if it can reveal a local land use authority's history of unnecessary restrictive limitations in its land use code or application process to religious organizations.[208]

(C.D. Cal. 2002). The plaintiff used RLUIPA's substantial burden provision when it sought to develop a more expansive property because it had outgrown its previous location. *See id.* at 1209, 1211–12. The court held that there was a substantial burden on plaintiff's religious exercise. *Id.* at 1224; *see also* Church of the Hills of Bedminster v. Twp. of Bedminster, No. Civ. 05-3332(SRC), 2006 WL 462674 (D.N.J. Feb. 24, 2006). The plaintiff church sought a variance for expansion under the defendant township's zoning regulations, which contained certain limitation rules on growth. *See id.* at *1. The plaintiff church withstood the township's motion to dismiss their RLUIPA substantial burden claim. *Id.* at *4–5. For a discussion of these cases with respect to expansion, see Daniel Dalton, *Expanding Existing Building and RLUIPA*, RLUIPA BLOG (Mar. 21, 2010, 9:07 PM), http://www.attorneysforlanduse.com/RLUIPABlog/?p=93.

204. *See* 42 U.S.C. § 2000cc(b)(3)(B) (2006).

205. *See id.*; *see also RMCC I*, 612 F. Supp. 2d 1163, 1176 (D. Colo. 2009) (noting the literal language of the provision is plural). Thus far, the Colorado district court's opinion has been the only court to place such emphasis on this plural requirement when analyzing the provision.

206. *See* 42 U.S.C. § 2000cc(a).

207. Specifically, the plaintiff in *West Pikeland* asserted that a limitation on the size of a building on its property to 5,000 square feet was unreasonable. *West Pikeland*, No. 09-cv-1626, 2010 WL 2635979, at *17 (E.D. Pa. June 25, 2010).

208. *RMCC II* is a prime example of this situation. *See RMCC II*, 613 F.3d 1229, 1238–39 (10th Cir. 2010).

Further, particularly if the case goes to trial, a religious claimant who brings both RLUIPA claims might have an advantage over a defendant municipality because a jury would have more legal theories under which to find fault. Also, bringing both claims might make RLUIPA claims easier for courts to adjudicate. For example, the Tenth Circuit in *RMCC II* sidestepped the issue of the substantial burden provision's constitutionality because the jury verdicts on RLUIPA's equal terms and unreasonable limitations claims were enough to support the district court's issuance of an injunction.[209]

In the end, whatever claim or claims parties assert in RLUIPA cases, one can see how a municipality's denial of a religious institution's application for expansion can present RLUIPA problems. Therefore, a local land use authority should consider potential RLUIPA claims when it is considering denying a religious institution's expansion application.[210] If not, the land use authority will likely find itself blindsided by an unexpected RLUIPA suit.

3. Unreasonable Limitations and Undue Discretion

Another lingering question is whether a land use board's practice of overly broad discretion within the land use review process amounts to a violation of the unreasonable limitations provision. In *Vision Church*, the Seventh Circuit reasoned that undue discretion of a zoning board does not violate the provision.[211] The *Vision Church* court supported this conclusion by reasoning that "the special use designation is substantially related to the municipal planning goals of limiting development, traffic and noise, and preserving open space; these goals, in turn, are reflected in the Village's Comprehensive Plan."[212] The court went on to say that secular institutions, not just religious organizations, also have to go through the special use permit process.[213] In the *Vision Church* court's view, a local board is entitled to broad discretion when equal requirements are placed on religious and non-religious entities, and even if zoning regulations grant undue discretion to the board, that discretion does not violate the provision.[214]

The *West Pikeland* court also suggested that the provision does not cover overly broad grants of discretion to a zoning board.[215] The court

---

209. Patricia Salkin, *10th Circuit Court of Appeals Upholds RLUIPA Jury Verdict in Mega Church Case*, LAW OF THE LAND (May 24, 2010), http://lawoftheland.wordpress.com/2010/05/24/10th-circuit-court-of-appeals-upholds-rluipa-jury-verdict-in-mega-church-case/.

210. *See RMCC II*, 613 F.3d at 1229; *see* Dalton, *supra* note 203 (discussing RLUIPA cases when a local government refuses expansion).

211. *Vision Church*, 468 F.3d 975, 991 (7th Cir. 2006).

212. *Id.* at 991.

213. *Id.*

214. *Id.*

215. *West Pikeland*, No. 09-cv-1626, 2010 WL 2635979, at *18 (E.D. Pa. June 25,

noted that overly broad grants of discretion are not addressed in the text of the statute, thus demonstrating that the unreasonable limitations provision might not apply to this type of discretion.[216] This interpretation is rational because neither the literal language nor the legislative history of RLUIPA discusses the relation of overly broad grants of discretion and unreasonable limits. However, the provision's language, which states that "[n]o government shall impose or implement a land use regulation that . . . unreasonably limits religious assemblies, institutions, or structures within a jurisdiction,"[217] is related to a local land use authority's discretionary measures. A land use authority might engage in discretionary measures to assess a special use application or to implement a zoning code. Depending on the circumstances, such discretionary measures are practices that could err on the side of being unreasonable and are indicative of the unreasonable limitations claims relied upon by on religious claimants.[218]

Overly broad grants of discretion to a local board might violate the unreasonable limitations provision within the context of a land use authority employing unnecessary restrictive practices. The factual record in *RMCC II* indicates a situation where an overly broad grant of discretion should constitute a violation of the provision. In *RMCC II*, the land use staff's actions arguably revealed a local land use authority operating with too much discretion and without procedural safeguards.[219] The county land use staff changed RMCC's report at a different stage than normal in the assessment process, and there was testimony of churches facing increasing restrictions in the past.[220] Broad grants of discretion should not be allowed, particularly when there are unequal requirements placed on religious institutions in the land use application process.

If the *Vision Church* court's analysis came to fruition—that there can be no violation of unreasonable limitations when procedural safeguards exist—Boulder County in *RMCC II* should have been given deference if its discretion was "narrowly circumscribed by the Village's Zoning Regulations."[221] Such a rationale of affording deference to a board would also be consistent with previous case precedent involving zoning

---

2010).

216. *Id. Compare West Pikeland*, 2010 WL 2635979, at *18 (implying that overly broad grants of discretion to a land use board do not apply with respect to the unreasonable limitations provision), *with* Salkin & Lavine, *supra* note 76, at 251 (reasoning that religious restrictions may be unreasonable based on the undue discretion of local officials when summarizing the unreasonable limitations provision).

217. 42 U.S.C. § 2000cc(b)(3)(B) (2006).

218. *See* Salkin & Lavine, *supra* note 76, at 251 ("There is some indication that religious restrictions may be unreasonable if they operate according to the unbridled discretion of local officials, but in such a case a plaintiff must do more than raise a facial challenge . . . .").

219. *See supra* Part II.C.

220. *See supra* Part II.C.

221. *Vision Church*, 468 F.3d 975, 990 (7th Cir. 2006).

decisions.[222] However, deference should not be afforded when a land use authority is going out of its way to impose restrictions on a particular religious claimant. This was reflected in the decisions of the district court and Tenth Circuit to allow RMCC to proceed with its expansion.[223] Thus, the unreasonable limitations provision should apply to the "unbridled discretion of local officials," particularly when no controls exist to ensure a fair land use application process or land use code.[224]

IV. FINDING COMMON GROUND

A. Land Use Regulation or Religious Freedom?

The unreasonable limitations provision, although only a minor part of RLUIPA, gives a religious institution the power to trump a reasonable land use regulation. Although local land use authorities are typically given discretion to require a special use application, variance, or exception, a religious assembly or institution might now use any increasing difficulty it faces in the review of these applications against that authority for a RLUIPA unreasonable limitations claim. This is precisely what happened in *RMCC II*, where a trend of increased difficulty in obtaining a permit rendered Boulder County in violation of RLUIPA.[225] Thus, when a land use authority limits a religious user's growth or location effort, and that religious user is equipped with evidence of the land use authority's past adverse treatment to churches, trouble is likely on the horizon for the land use authority.

The situation described above is ripe for a RLUIPA suit, particularly a RLUIPA unreasonable limitations suit. A rise in the number of lawsuits like these will result in an increasing threat to certain land use goals. For instance, this trend of churches using RLUIPA to overcome reasonable land use standards can and will likely lead to the frustration of comprehensive plans.[226] Boulder County had a comprehensive plan of limiting urban sprawl and preserving open space.[227] Unfortunately for Boulder County, RMCC was able to use RLUIPA to overcome that plan.[228] RMCC

222. *See generally* City of Los Angeles v. Alameda Books, Inc., 535 U.S. 425 (2002) (upholding a city ordinance that prohibited more than one adult entertainment business in the same building); City of Renton v. Playtime Theatres, Inc., 475 U.S. 41 (1986) (holding that the city of Renton was entitled to rely on the experience of Seattle and other cities in enacting its adult theater zoning ordinance).

223. *See supra* notes 146–154 and accompanying text.

224. *See* Salkin & Lavine, *supra* note 76, at 251.

225. *RMCC II*, 613 F.3d 1229, 1238 (10th Cir. 2010).

226. *See supra* Part III.B.

227. *RMCC II*, 613 F.3d at 1233.

228. *Id.* at 1240.

specifically used RLUIPA when it was denied an expansion opportunity.[229] RMCC played a kind of legal lottery, using numerous RLUIPA provisions, to throw out many claims with the hope that one would stick. The church only needed to succeed on a few RLUIPA provisions or, in fact, one to obtain an injunction on the County's denial of their proposed development.[230] *RMCC II* should put municipalities on notice that churches encountering similar circumstances could prevail on their own RLUIPA claims.

This demonstrates an inherent problem with RLUIPA because it might equip religious institutions with too much capability of frustrating comprehensive plans and, more particularly, local land use rules and regulations. Comprehensive plans set out goals for the betterment of a jurisdiction's citizenry.[231] Generally, a goal in a comprehensive plan that aims to limit development correlates to the preservation of open space, traffic control, or building density.[232] These effects attained from limiting development help provide for less traffic congestion, which relates to a higher quality of life for the community.[233] However, with an increasing awareness of RLUIPA, local land use authorities now might face the prospect of unattainable comprehensive plans. With religious institutions having the power of development or expansion, land use authorities will likely encounter difficulties in their planning goals, potentially resulting in the detriment of community interests.[234]

However, this frustration of land use regulations, or at least the risk of it, is sometimes necessary to resolve the "covert" discrimination that has and is still taking place in the land use context.[235] In accordance with RLUIPA, a local land use authority needs to protect a person's religious exercise, particularly in land use decisions.[236] This is demonstrated by RLUIPA's hearing record. The joint statement from Senators Hatch and Kennedy explains that "[c]hurches in general, and new, small, or unfamiliar churches in particular, are frequently discriminated against on the face of zoning codes and also in the highly individualized and discretionary

---

229. *Id.* at 1235.

230. Specifically, RMCC's injunction was upheld based on the equal terms and unreasonable limitations provisions of RLUIPA. *Id.* at 1240.

231. *See, e.g.*, Smith v. Zoning Bd. of Appeals of Town of Greenwich, 629 A.2d 1089, 1098 (Conn. 1993) ("The purpose of [a comprehensive town] plan is to set forth the most desirable use of land and an overall plan for the town.").

232. *See supra* note 26; *see, e.g.*, *RMCC II*, 613 F.3d at 1233.

233. *See supra* notes 167–173 and accompanying text.

234. *See generally* Weinstein, *supra* note 31, at 4 (explaining that churches often hold events late into the night and expand their activities throughout the day, increasing neighbor concerns about the effects of such activities on their property values).

235. 146 CONG. REC. 16,698, 16,699 (2000) (joint statement of Sens. Hatch and Kennedy).

236. *See supra* Part I.B. (discussing the need for the enactment of the statute).

processes of land use regulation."[237] The hearing record further explains that this religious discrimination takes place in the form of zoning codes by excluding churches entirely or through a zoning board's individualized assessment process.[238] RLUIPA has a broad definition of "religious exercise" for a reason.[239] The definition goes to correcting past and current discrimination taking place in the land use context,[240] thereby corroborating existing First Amendment values.

The tension between land use goals and religious freedom is difficult to balance and will likely grow. When the scale tips too far in one direction or the other, drastic long-term consequences may result because religious freedom or planning goals are frustrated. Land use regulation and the attainment of land use goals are important, but ending religious discrimination in the land use context is also imperative, as it supports the free exercise of religion. Since our nation has become accustomed to broad free religious exercise protection, comprehensive plans and local land use implementation to meet those plans will likely need to make way for religious freedom as the practice of religion rises. This means the prevalence of more RLUIPA claims, further complicating matters for land use authorities, who now must learn to operate with a consciousness of RLUIPA.

B. A Future Perspective

New ideas have emerged to curb the tension between land use authorities and religious entities. For example, one recently introduced concept relates to the operation of a mixed-use building as a business that leases spaces to a church.[241] Pastor Jeff Vanderstelt of Soma Church in

---

237. 146 CONG. REC. 16,698, 16,698 (joint statement of Sens. Hatch and Kennedy).

238. *Id.*

239. *See* 42 U.S.C. § 2000cc-5(7)(A) (2006).

240. *See* 146 CONG. REC. 16,698, 16,698 (joint statement of Sens. Hatch and Kennedy).

241. Mark Bergin, *New Approach to Church Real Estate Seeks to Dispel Civic Resentment Over Lost Tax Revenue*, WORLD MAGAZINE, June 9, 2007, *available at* http://www.worldmag.com/articles/13031. To illustrate the resentment between churches and land use authorities, consider this example:

> Stafford, Texas, a Houston suburb of fewer than 20,000 people, has 51 churches. On Sunday mornings, the town swells with commuting church members, clogging roadways and public spaces with congestion that civic leaders can't afford to solve. The tax-exempt houses of worship occupy so much of the city's seven square miles that tax revenue can barely cover police, fire, and schools, never mind new development. The predicament has pushed government officials to explore legal pathways to block church growth.

Id.

Tacoma, Washington, purchased a property and formed a for-profit LLC independent of the church.[242] Vanderstelt's company leases space to commercial businesses while also offering a highly discounted lease rate to Soma Church.[243] The company uses extra money derived from this source of revenue to provide salaries for the church staff.[244] Moreover, Vanderstelt has paid all the property tax on this building.[245] Not only does this mixed-use approach provide an innovative source of revenue to the city, but it provides for an "open house philosophy" that integrates Soma church members into the community.[246] Since the building is owned by a business, there is no hesitancy by schools or other entities in using the facility.

It is this type of thinking that would advantage land use authorities as well as religious entities. Vanderstelt's idea of mixed-use development encourages a balance between churches and land use regulation; you have a business operating a church that provides revenue to a city, relieving any tension that can result from lost property taxes. Further, community integration results because people of different religious faiths do not feel subjected to one faith while in that mixed-use building, as it is not actually owned by the church.[247] Although this concept might require outside investment, such an approach is a step in the right direction. Yet another caveat with this approach is that it might face difficulty in its application with regard to megachurches. A much more large-scale expenditure is required by a willing investor in such a context. Moreover, Vanderstelt's approach might not cure the traffic and congestion problems that a megachurch creates, despite the approach involving mixed-use building owners paying money to a community's tax base. In some instances, the benefits of additional money to the tax base might be outweighed by the traffic and congestion problems posed by the existence of a megachurch. Thus, megachurches are a potential problem for this approach.

These types of churches also raise questions about the effectiveness of RLUIPA.[248] In light of RLUIPA's unreasonable limitations language that

---

242. *Id.* Vanderstelt was able to acquire financing from investors at eight percent interest. *Id.*

243. *Id.*

244. *Id.*

245. *Id.*

246. *Id.* Vanderstelt's idea was created from his past experiences at megachurches. *Id.* As a youth pastor at a church of 20,000 people in South Barrington, Illinois, he questioned a $90 million building project that exhausted a community's resources. *Id.* He recalled, "We had to hire a whole PR department to deal with all the problems the building caused . . . . The city was saying, 'We have to pay for the traffic problems that you've created, but you're not paying to the tax base to take care of it.'" *Id.*

247. *Id.*

248. *See also* Lillian Kwon, *Report Tracks Megachurch Growth, Changes*, THE CHRISTIAN POST (Sept. 13, 2008, 9:16 AM), http://www.christianpost.com/article/20080913/report-tracks-megachurch-growth-changes (noting that "church planting has grown among megachurches from 68 percent in 2000 . . . to 77 percent in 2008"). *See generally* Scott

government may not unreasonably limit religious assemblies, institutions, or structures within a jurisdiction,[249] the assessment of what actual limitations exist on religious entities is significant. It is important to understand what are and are not reasonable limitations within the realm of RLUIPA.[250]

A particular problem regarding megachurches involves the intertwining of secular with religious land uses.[251] If a religious entity is restricted to locate within a jurisdiction because its additional amenities will distress a community, RLUIPA's unreasonable limitations provision might provide recourse to that religious entity in its location effort, assuming there has been more than one incidence of unreasonable limits.[252] In accordance with this provision, the religious claimant could allege a violation of it because it had been unfairly restricted from a jurisdiction. Further, the claimant could potentially assert that its secular uses deserve protection as "religious exercise"[253] under RLUIPA because they advance a religious purpose.[254] This illustration reflects a limitations problem in RLUIPA because of its failure to address secular land use by religious entities.

This limitations problem has potentially adverse effects for communities if RLUIPA claimants begin to succeed in intertwining secular and religious land use.[255] Additional secular amenities can produce increased noise, traffic and pollution.[256] Further, an inherent unfairness exists when a secular use coupled with a religious use can be protected under RLUIPA, but a secular use alone would be subject to additional restrictions in the land use context.[257] This problem exemplifies the need for

---

Thumma, *Exploring the Megachurch Phenomena: Their Characteristics and Cultural Context*, HARTFORD INSTITUTE FOR RELIGION RESEARCH, http:// hirr.hartsem.edu/ bookshelf/thumma_article2.html (last visited Mar. 18, 2011) (demonstrating that megachurches are experiencing rapid church growth).

249. 42 U.S.C. § 2000cc(b)(3)(B) (2006).

250. The *Konikov* court reasoned that "variances, special permits or exceptions, hardship approval, or other relief provisions in land use regulations" are reasonable limitations. Konikov v. Orange Cnty., Florida, 302 F. Supp. 2d 1328, 1346 (M.D. Fla. 2004), *aff'd in part, rev'd in part*, 410 F.3d 1317 (11th Cir. 2005) (citing 146 CONG. REC. S7,774, 7,776 (daily ed. July 27, 2000) (joint statement of Sens. Hatch and Kennedy)).

251. Welch, *supra* note 27, at 256–58, 260.

252. *See RMCC I*, 612 F. Supp. 2d 1163, 1176–77 (D. Colo. 2009).

253. The term "'religious exercise' includes any exercise of religion, whether or not compelled by, or central to, a system of religious belief." 42 U.S.C. § 2000cc-5(7)(A) (2006).

254. Welch, *supra* note 27, at 260.

255. *Id.* at 258.

256. *Id.*

257. *See generally* Salkin & Lavine, *supra* note 76 ("[T]he reality is that RLUIPA has had the effect of burdening local governments and providing religious groups with an unfair advantage not available to other applicants.").

federal legislation, such as a RLUIPA amendment, to help define appropriate limitations on religious entities.[258]

With respect to the vagueness of RLUIPA's provisions, the United States Supreme Court could resolve conflicting views among circuits by ruling on the validity of RLUIPA's land use provisions.[259] Although the U.S. Supreme Court denied Boulder County's petition for certiorari, the issue of the validity of RLUIPA's land use provisions is bound to come before the Court again in the future.[260] Nonetheless, Congress should resolve the vagueness issues that RLUIPA and its legislative history present by enacting new legislation to cure some of these defects. RLUIPA's substantial burden and the unreasonable limitations provisions are starting points in which future judicial or legislative insight is necessary. Otherwise, conflicting views of RLUIPA's language will remain, leaving courts to employ their own tests under RLUIPA's provisions.

These particular problems highlight RLUIPA's ineffectiveness, but they do not necessarily warrant an invalidation of RLUIPA's land use provisions altogether. Certainly, this would be a drastic measure considering the effort and events leading up to the statute's enactment in addition to the new land use regulations that would have to be enacted to ensure the protection of religious freedom.[261] The broad First Amendment protection that RLUIPA provides is necessary, as demonstrated by recent

---

258. Sara C. Galvan, *Beyond Worship: The Religious Land Use and Institutionalized Persons Act of 2000 and Religious Institutions' Auxiliary Uses*, 24 YALE L. & POL'Y REV. 207, 210 (2006) ("More and more, megachurches desperate for larger spaces are achieving their expansionist goals with the help of RLUIPA. And there is every reason to think they will continue to do so."). A Boulder County attorney stated that "the reach of RLUIPA 'will expand to the point where religious institutions are effectively dictating their own land[]use regulations.'" Salvin & Lavine, *supra* note 76, at 254 (citing Diana B. Henriques, *Religion Trumps Regulation as Legal Exemptions Grow*, N.Y. TIMES, Oct. 8, 2006, at A1, *available at* http://www.nytimes.com/2006/10/08/business/08religious.html?_r=1&oref=slogin).

259. *See* MacLeod, *supra* note 6, at 54 ("Courts have struggled to articulate a consistent standard for applying ["substantial burden,"] and there is at present no general consensus on the best construction, as there is with the terms 'land use regulation' and 'religious exercise.'"); *see also* Evans, *supra* note 116 ("Boulder County isn't the only place where RLUIPA is being challenged. Federal courts are coming to different conclusions about RLUIPA's terms and meaning. Even if the county decides not to appeal the issue to the U.S. Supreme Court, someone will, and soon.").

260. The Boulder County Board of Commissioners voted and approved to ask the U.S. Supreme Court to weigh in on this case. Laura Snider, *Boulder County Will Ask U.S. Supreme Court to Consider Rocky Mountain Christian Church Case*, BOULDER DAILY CAMERA (Sep. 30, 2010, 11:06 AM), http://www.dailycamera.com/news/ci_16212241. The County's petition for certiorari was denied on January 10, 2011. Bd. of Cnty. Comm'rs of Boulder County, Colo. v. Rocky Mountain Christian Church, 131 S. Ct. 978, 978 (2011).

261. *See supra* Part II.A (discussing events leading up to RLUIPA's enactment and the need for the legislation).

cases.[262] It is essential that a federal statute such as RLUIPA exist where religious claimants can seek recourse when more than necessary limitations burden these particular claimants. The aforementioned secular land use problem posed by megachurches and RLUIPA's vague provisions should not be enough to warrant an invalidation of the statute.[263] This secular land use problem could likely be cured by supplemental federal legislation addressing appropriate limitations on religious entities, or more specifically, secular land use by religious entities. If RLUIPA were invalidated, other federal legislation would be required, leaving some of the underlying values of RLUIPA intact while also aiming to cure some of these unanticipated concerns.

Another proposed solution to RLUIPA's inadequacies is to return land use regulation to state and local governments.[264] This approach likely has more disadvantages than advantages.[265] Although our nation is accustomed to deferring to state and local governments in land use matters,[266] religious freedom mandates federal oversight. Local land use authorities might be in a better position to tailor land use regulations to a specific dispute,[267] but it is important to keep these local land use authorities in check to eliminate any unfairness religious users experience in the land use context.[268] However, an advantage of a return to local land use regulation is that municipalities, in drafting land use regulations and goals, do not have to go too far in the protection of religious use because of the prospect of a RLUIPA suit.[269] This result, on the other hand, is an inescapable consequence of RLUIPA. Land use authorities will have to learn to adopt more accommodating zoning codes and fair land use procedures until some resolution emerges.

---

262. *See, e.g.*, *RMCC II*, 613 F.3d 1229 (10th Cir. 2010). Specifically, RMCC was arguably mistreated during the assessment of its 2004 special use application. *See supra* Part II.C. RMCC had recourse in RLUIPA, which exposed a suspect land use process by Boulder County in RMCC's 2004 application. *See RMCC II*, 613 F.3d at 1234.

263. *See, e.g.*, Salkin & Lavine, *supra* note 76, at 256 ("With the benefit of hindsight, it is now fairly clear that RLUIPA's overinclusiveness has done more harm than good."). *But see* Welch, *supra* note 27, at 256 (arguing for an invalidation of RLUIPA, in part, due to this secular land use problem).

264. Welch, *supra* note 27, at 294.

265. *But see* Marci A. Hamilton, *Federalism and the Public Good: The True Story Behind the Religious Land Use and Institutionalized Persons Act*, 78 IND. L.J. 311, 320–23 (discussing reasons why state and local governance is preferable to federal governance); Salkin & Lavine, *supra* note 76, at 254–55 (illustrating the impact RLUIPA is having on state and local governments).

266. Welch, *supra* note 27, at 294–95.

267. *See id.* at 295.

268. *But see* Salkin & Lavine, *supra* note 76, at 258 (explaining that "states were not lacking in statutory religious protections" before RLUIPA's passage).

269. Welch, *supra* note 27, at 297.

V. CONCLUSION

As of now, the solution for the inadequacies of RLUIPA's land use provisions falls with the courts to evaluate claims on a case-by-case basis. With respect to cases involving RLUIPA's unreasonable limitations provision, this approach would be consistent with RLUIPA's legislative history instructing courts to make determinations "in light of all facts."[270] Case precedent on the unreasonable limitations provision has produced rulings that create uncertainty for land use authorities[271] who more and more should be focusing on smart growth as the nation's population grows and brings more development.

RLUIPA's unreasonable limitations cases, however, have also revealed certain lessons. Land use authorities should learn to pay particular attention to this provision when denying an expansion opportunity, particularly if they have created an unnecessarily restrictive environment for religious claimants. Further, an overly broad grant of discretion to a zoning board is likely not violative of the provision, unless a local land use authority has gone out of its way to place restrictions or obstacles on particular religious claimants. *RMCC II* is emblematic of circumstances where a local land use authority's undue discretion led to a violation of the provision.[272] Finally, the provision applies when municipalities broadly limit where religious entities can locate, an application consistent with case law and RLUIPA's legislative history. However, this Comment has shown that other evidence is considered when assessing whether a violation of the provision has occurred.[273]

RLUIPA has created tension between religious entities and land use authorities. While religious exercise is important, religious developers need to cooperate with land use authorities to achieve sustainable growth in the future. Vanderstedlt's mixed-use approach, which allows religious entities to become integrated into a community while also being part of a building whose owner is paying property taxes, is at least a start.[274] However, the approach does have its flaws, as illustrated by the type of investment and congestion concerns a megachurch presents. Further, the rise of megachurches with their secular amenities exposes an absence of limitations issue in RLUIPA.[275]

---

270. 146 CONG. REC. 19,123 (2000) (statement of Rep. Canady on RLUIPA).

271. *See supra* Part III.

272. *RMCC II*, 613 F.3d 1229, 1229 (10th Cir. 2010).

273. *See* Part III.C.

274. *See supra* notes 241–246 and accompanying text.

275. Jennifer S. Evans-Cowley & Kenneth Pearlman, *Six Flags Over Jesus: RLUIPA, Megachurches, and Zoning*, 21 TUL. ENVTL. L.J. 203, 208 ("Lawsuits and public disputes over zoning of megachurches are increasing in regularity throughout the country.") (citing Mark Bergin, *Bullied Pulpits*, WORLD MAGAZINE, May 20, 2006, *available at* http://www.worldmag.com/articles/11864).

In light of these particular problems, RLUIPA might have created "more harm than good."[276] Such arguments beg the question of whether RLUIPA's land use provisions have met their end. Depending on future religious building development and its resulting problems, this notion may be realized. However, it is important to remember that when enacting any new legislation in the area of religious land use, federal oversight is essential. Despite RLUIPA's flaws, the underlying aims of the statute will need to remain intact to maintain the free exercise of religion that our Constitution requires and to which our society has become accustomed.[277]

276. Salkin & Lavine, *supra* note 76, at 256.

277. Quoting Thomas Jefferson, one blogger writes, "I would rather be exposed to the inconveniences attending too much liberty, than those attending too small a degree of it." Charlie Danaher, *What's Wrong with Letting Churches Expand?* BOULDER DAILY CAMERA (Aug. 1, 2010, 1:00 AM), http://www.dailycamera.com/ci_15642582?IADID=Search-www.dailycamera.com-www.dailycamera.com.

CONSTITUTIONAL LAW—FIFTH AMENDMENT—HONING THE PROTECTIONS OFFERED BY *MIRANDA* THROUGH SUPREME COURT BRIGHT-LINE RULEMAKING

Maryland v. Shatzer, 130 S. Ct. 1213 (2010).

I. INTRODUCTION

On August 7, 2003, Michael Shatzer, Sr., the defendant, was questioned regarding the alleged sexual child abuse of his then three-year-old son.[1] At the time of the inquiry, Shatzer was serving time at the Maryland Correctional Institute-Hagerstown.[2] Prior to conducting the interrogation, Detective Blankenship read Shatzer his *Miranda* rights, which he promptly waived.[3] However, since he misunderstood the reason for Blankenship's presence, Shatzer "declined to speak without an attorney."[4] Blankenship properly ended the interrogation and "Shatzer was released back into the general prison population."[5]

Two-and-a-half years later, the discovery of new information led Detective Paul Hoover to reopen the investigation.[6] On March 2, 2006, Hoover interviewed Shatzer and read him his *Miranda* rights; Shatzer again waived his rights via a written form.[7] Shatzer cooperated with the investigation and "[a]t no point during the interrogation did [he] request to speak with an attorney or refer to his prior refusal to answer questions without one."[8] During the questioning, Shatzer alluded to some criminal activity[9] and agreed to a polygraph test.[10] Five days later, Shatzer failed the polygraph test, made incriminating remarks,[11] and requested counsel.[12]

---

1. Maryland v. Shatzer, 130 S. Ct. 1213, 1217 (2010). The allegations arose when the Child Advocacy Center of the Criminal Investigation Division of the Hagerstown Police Department conducted a preliminary investigation. *Id.* Detective Shane Blankenship, an employee of the Hagerstown Police Department, questioned Shatzer. *Id.*

2. *Id.* Shatzer's incarceration at the time of questioning was due to an "unrelated child-sexual-abuse offense." *Id.*

3. *Id.* Shatzer waived his rights through written consent. *Id.*

4. *Id.* Shatzer believed that "Blankenship was an attorney there to discuss the prior crime for which he was incarcerated." *Id.*

5. *Id.* After which, "Blankenship closed the investigation." *Id.*

6. *Id.* New evidence arose from interviews with the victim, now eight years old, when the same social worker unearthed new information. *Id.* at 1218.

7. *Id.* Hoover conducted the investigation in a maintenance room at the correctional facility where Shatzer was incarcerated. *Id.*

8. *Id.*

9. *Id.* Shatzer confessed to "masturbating in front of his son from a distance of less than three feet" but "denied ordering his son to perform fellatio on him." *Id.*

At trial, he moved to suppress his incriminating statements from March 2006 based on *Edwards* protections.[13] The court denied his motion, holding that "[t]he *Edwards* protections did not apply . . . because Shatzer had experienced a break in custody for *Miranda* purposes between the 2003 and 2006 interrogations."[14] Shatzer was subsequently convicted based on the evidence used during a bench trial.[15] The Court of Appeals of Maryland reversed and remanded the trial court's decision, holding that "the passage of time *alone* is insufficient to expire the protections afforded by *Edwards*."[16] The court further argued that it need not address the legitimacy of a "break in custody" exception to *Edwards* because Shatzer never experienced such a break, as he was in continuous custody between interrogations for the same offense.[17] On certiorari to the United States Supreme Court, *held*, reversed and remanded.[18] If an individual who has previously asserted his *Miranda* right to counsel has been out of custody for fourteen days, the protections offered by *Edwards* do not apply.[19] Further, a break in custody is defined by the point at which the "inherently compelling pressures" of the interrogation cease and the accused "return[s] to his normal life," regardless of if one's "normal life" exists in prison.[20] *Maryland v. Shatzer*, 130 S. Ct. 1213 (2010).

---

10. *Id.*

11. *Id.* Shatzer stated, "'I didn't force him. I didn't force him.'" *Id.* (quoting Shatzer v. State, 954 A.2d 1118, 1121 (Md. 2008)).

12. *Id.* The investigator ended the questioning following Shatzer's request for counsel. *Id.* After these remarks, he was formally charged with "second-degree sexual offense, sexual child abuse, second-degree assault, and contributing to conditions rendering a child in need of assistance." *Id.*

13. *Id.* The "*Edwards* protections" refer to *Edwards v. Arizona*, 451 U.S. 477 (1981), where the Court held that "it is inconsistent with *Miranda* and its progeny for the authorities . . . to reinterrogate an accused in custody if he has clearly asserted his right to counsel." *Id.* at 485. Interrogation must cease until counsel is made available to the accused "unless the accused himself initiates further communication, exchanges, or conversations with the police." *Id.* at 484–85.

14. *Shatzer*, 130 S. Ct. at 1218.

15. *Id.*

16. *Shatzer*, 954 A.2d at 1131, 1138 (alteration in original).

17. *Id.* at 1133. The court notes that the break in custody exception recognized by other courts pertained to individuals released back into the general populace. *Id.*

18. *Shatzer*, 130 S. Ct. at 1227. The Court granted certiorari at Maryland v. Shatzer, 129 S. Ct. 1043 (2009).

19. *Shatzer*, 130 S. Ct. at 1223.

20. *Id.* at 1225.

II. DEVELOPMENT OF THE FIFTH AMENDMENT'S SELF-INCRIMINATION CLAUSE

The Fifth Amendment to the United States Constitution provides that "[n]o person . . . shall be compelled in any criminal case to be a witness against himself."[21] Scholars note that the self-incrimination clause and its protections were historically hard to define and that even today, "the clause continues to confound and confuse."[22] Through varying interpretations and implementations, the clause has managed to mold and influence modern-day police procedures.[23] In *Shatzer*, two issues induced the Court to grant certiorari: first, whether a break in custody ended Fifth Amendment protections offered by *Miranda*;[24] and second, whether release back into a general prison population constituted a break in *Miranda* custody.[25] The United States Supreme Court sought to clarify the extent and nature of the self-incrimination clause through a holding which offered a bright-line rule for courts and police personnel to follow in order to resolve these issues.[26] The Court first held that after invoking one's *Miranda* rights, the resulting *Edwards* protections last for fourteen days, after which the police may conduct a second interrogation without violating those rights.[27] Second, the Court held that a prisoner's release back into a general prison population did constitute a break in custody for *Miranda* purposes.[28]

A. Pre-Miranda Protections

Historically, the Supreme Court has sought to protect a defendant's privilege against self-incrimination during testimony and questioning.[29] As a confession is the ultimate self-incrimination, the Supreme Court sought to clarify the Fifth Amendment's self-incrimination provision throughout the

---

21. U.S. CONST. amend. V. This provision is commonly referred to as the Self-Incrimination Clause of the U.S. Constitution. AKHIL REED AMAR, THE CONSTITUTION AND CRIMINAL PROCEDURE: FIRST PRINCIPLES 46 (1997).

22. AMAR, *supra* note 21, at 46.

23. *Id.* at 48. Amar notes that the clause has "narrow[ed]" law enforcement's fact gathering and interactions with individuals. *Id.* He concludes that this has pushed the officials to seek other methods of obtaining information without entering the sphere of the Fifth Amendment's protections. *Id.*

24. *Shatzer*, 130 S. Ct. at 1220.

25. *Id.* at 1224.

26. *Id.* at 1223–24. The Court states, "And when it is determined that the defendant pleading *Edwards* has been out of custody for two weeks before the contested interrogation, the court is spared the fact-intensive inquiry into whether he ever, anywhere, asserted his *Miranda* right to counsel." *Id.*

27. *Id.* at 1223.

28. *Id.* at 1225.

29. *See* Counselman v. Hitchcock, 142 U.S. 547 (1892).

early twentieth century.[30] In the 1950s, the Court attempted to institute tests to determine the voluntariness of an individual's confession.[31] The lack of consistency and degree of subjectivity required in applying these tests left a "growing dissatisfaction with the voluntariness standard" the Court then employed.[32] The Court attempted to battle the voluntariness standard's ambiguity with the "objective constitutional guidelines" it put forth in *Miranda*.[33]

B. Miranda Protections

The Supreme Court set clear goals and objectives with its decision in *Miranda v. Arizona*.[34] The majority was careful to clarify that "our holding is not an innovation in our jurisprudence, but is an application of principles long recognized and applied in other settings."[35] The Court first analyzed the "admissibility of statements obtained from a defendant questioned while in custody or otherwise deprived of his freedom of action in any significant way"[36] by describing coercive custodial interrogations.[37] The Court's main

---

30. *See* Mary A. Crossley, Note, Miranda *and the State Constitution: State Courts Take a Stand*, 39 VAND. L. REV. 1693, 1704 (1986).

31. *Id.* at 1705. The Court implemented the "'forced' confession exclusion" which upheld the Court's earlier determination that, if a confession was deemed coerced, then other evidence would have to be produced to support that the confession was, in fact, voluntary. *Id.* The second test, deemed the "police methods" test, sought to examine the legitimacy of the police procedural process through which the confession was obtained. *Id.* These tests were considered when the "totality of each case's circumstances" were considered. *Id.* It should also be noted that the court considered many other factors, including but not limited to "the intelligence, physical health and emotional characteristics of the particular suspect" and "his age, education and prior criminal record" when evaluating the voluntariness of the individual's confession. Yale Kamisar, *On the Fortieth Anniversary of the* Miranda *Case: Why We Needed It, How We Got It—And What Happened to It*, 5 OHIO ST. J. CRIM. L. 163, 163 (2007).

32. Crossley, *supra* note 30, at 1705. This "dissatisfaction" arose from "the requirement of a case-by-case analysis" and the "unpredictability of outcomes [that] left police with no specific rules for interrogation practices." *Id.*

33. *Id.* at 1706.

34. Miranda v. Arizona, 384 U.S. 436 (1966). The Court "granted certiorari in these cases . . . in order further to explore some facets of the problems, thus exposed, of applying the privilege against self-incrimination to in-custody interrogation, and to give concrete constitutional guidelines for law enforcement agencies and courts to follow." *Id.* at 441–42.

35. *Id.* at 442. The Court is alluding to the rights set forth and provided for by the Fifth Amendment of the U.S. Constitution. More specifically it notes the Self-Incrimination Clause and the right to counsel. *Id.*

36. *Id.* at 445.

37. *Id.* The Court describes examples of various in-custody interrogations where physical abuse is displayed by law enforcement. *Id.* at 446. The Court notes, however, that the "modern practice of in-custody interrogation is psychologically rather than physically oriented." *Id.* at 448. The psychological effects that police seek are elicited through a private,

concern was this intimidating, police-dominated atmosphere and the "evils it can bring."[38]

In *Miranda*, the Court described the privilege that the Self-Incrimination Clause offers to citizens as "the respect a government—state or federal—must accord to the dignity and integrity of its citizens."[39] In its analysis, the Court attempted to protect the individual's Fifth Amendment rights when subjected to the coercive, police-dominated atmosphere in which custodial interrogations take place.[40] These rights include the right to remain silent[41] and the right to counsel.[42] Upon arrest and interrogation, suspects must be informed "clear[ly] and unequivocal[ly]" of their rights;[43] otherwise, all admissions made during the interrogation are inadmissible.[44] However, the Court was careful to not discount all admissions of guilt: "Confessions remain a proper element in law enforcement."[45] In *Miranda*, the Court effectively adopted these procedural safeguards and maintained that it was necessary "unless other fully effective means are adopted."[46]

C. Exceptions to and Extensions of Miranda

As the composition of the Supreme Court shifted over the next forty years, *Miranda*'s holding experienced moments of both protective

---

unfamiliar location with police officers who employ a tactful "aura of confidence in [the suspect's] guilt," coupled with patient, persistent, and relentless questioning. *Id.* at 450, 455.

38. *Id.* at 456.

39. *Id.* at 460.

40. *Id.* at 467, 478–79. These measures include explicitly informing the person of the right to remain silent and the right to counsel through hire or appointment. *Id.* at 468, 471.

41. *Id.* at 467–68. The Court states that this right must "be accompanied by the explanation that anything said can and will be used against the individual in court." *Id.* at 469.

42. *Id.* at 470. The Court had previously explained the importance of counsel in their ability to allow the defendant to "tell his story without fear, effectively, and in a way that eliminates the evils in the interrogation process." *Id.* at 466.

43. *Id.* at 467–68.

44. *Id.* at 479.

45. *Id.* at 478. The Court holds that "[v]olunteered statements of any kind are not barred by the Fifth Amendment and their admissibility is not affected by our holding today." *Id.*

46. *Id.* at 479.

expansion and restriction.[47] Subsequent holdings attempted to further define and interpret many of the key terms and ideas that *Miranda* set forth.[48]

While *Miranda* found much support in the Warren Court,[49] it failed to find similar sentiments in the successor Burger Court.[50] With the Court's new composition came new considerations as the Burger Court "quickly set clear limits on *Miranda*'s potentially far-reaching implications."[51] The Burger Court passed down several holdings that questioned *Miranda*'s applicability to various fact patterns[52] and its categorization as a "mere safeguard[]" and not a "constitutional right."[53] For example, in *Harris v. New York*, the Burger Court found that, while incriminating statements obtained without *Miranda* warnings could not be introduced as evidence, they could be used to impeach the defendant when testifying.[54] The Court stated, "The shield provided by *Miranda* cannot be perverted into a license to use perjury by way of a defense, free from the risk of confrontation with prior inconsistent utterances."[55] In his dissent, Justice Brennan found the majority's decision to "seriously undermine the achievement of [*Miranda*'s] objective" of "deterring police practice in disregard of the Constitution."[56] While this holding limited the protections of *Miranda*, the blatant disagreement within the Court is noteworthy, as it demonstrated the fickle relationship the Court experienced—and continues to experience—with *Miranda*.[57]

In 1981, the Court decided *Edwards v. Arizona*,[58] a decision that its successors would eventually call "a second layer of prophylaxis for the *Miranda* right to counsel."[59] To establish a valid waiver of one's Fifth and

---

47. Crossley, *supra* note 30, at 1708–09. Crossley comments that there have been varying interpretations, especially that of the Burger Court. *Id.* at 1709. She notes that some may react to the Burger Court by saying that it concentrated on the unimportant, trivial aspects of *Miranda*, while others may find it closely followed its doctrine. *Id.* Crossley, however, is clear to note that "the Burger Court quickly set clear limits on *Miranda*'s potentially far-reaching implications." *Id.* at 1709.

48. Kamisar, *supra* note 31, at 180. "Key terms" include custody, custodial interrogation, waiver, etc. *Id.*

49. Crossley, *supra* note 30, at 1708. The Warren Court decided *Miranda*. *Id.*

50. *Id.* at 1709.

51. *Id.*

52. *Id.* at 1709–10; *see, e.g.*, Harris v. New York, 401 U.S. 222 (1971) (admitting statements that were inadmissible under *Miranda* to be used in impeachment).

53. Crossley, *supra* note 30, at 1711; *see, e.g.*, Michigan v. Tucker, 417 U.S. 433 (1974) (admitting statements made before the *Miranda* decision and given after the defendant asserted his privilege against self-incrimination but not his right to counsel).

54. *Harris*, 401 U.S. at 226.

55. *Id.*

56. *Id.* at 232 (Brennan, J., dissenting).

57. *See id.* (Brennan, J., dissenting).

58. Edwards v. Arizona, 451 U.S. 477 (1981).

59. McNeil v. Wisconsin, 501 U.S. 171, 176 (1991).

Sixth Amendment rights, *Miranda* held that a waiver may not be interpreted from silence;[60] rather, a waiver of one's rights must be one of a "high standard,"[61] and they must be waived "competently and intelligently."[62] The Court in *Edwards* confirmed *Miranda*'s holding and further expanded its protections.[63] While *Miranda* required that law enforcement end the interrogation once the accused invokes the right to remain silent or to seek the advice of counsel,[64] *Edwards* further held that "when an accused has invoked his right to have counsel present during custodial interrogation, a valid waiver of that right cannot be established by showing only that he responded to further police-initiated custodial interrogation."[65] In *Edwards*, the Court found the defendant's incriminating statement inadmissible because the incriminating "statement, made without having had access to counsel, did not amount to a valid waiver and hence was inadmissible" as the defendant had invoked his right to counsel the night before and had not volunteered information, but was provoked by police questioning.[66]

Shortly after the *Edwards* decision, the Supreme Court again limited *Miranda*'s scope. In *New York v. Quarles*, the Court held that "there is a 'public safety' exception to the requirement that *Miranda* warnings be given before a suspect's answers may be admitted into evidence."[67] The Court concluded that the need to obtain information from a suspect to avert potentially hazardous situations "outweighs the need for the prophylactic

60. Miranda v. Arizona, 384 U.S. 436, 475 (1966).

61. *Id.* (citing Johnson v. Zerbst, 304 U.S. 458 (1938)).

62. *Zerbst*, 304 U.S. at 468. The Court found that, in assessing the legitimacy of a waiver, it "depend[s] . . . upon the particular facts and circumstances surrounding that case, including the background, experience, and conduct of the accused." *Id.* at 464.

63. *Edwards*, 451 U.S. at 484.

64. *Miranda*, 384 U.S. at 474.

65. *Edwards*, 451 U.S. at 484. The Court held that the individual should not be interrogated further until counsel arrives. If, however, the individual "initiates further communication, exchanges, or conversations with the police," he may be questioned further. *Id.* at 484–85.

66. *Id.* at 487. The defendant invoked his right to counsel during his first interrogation. The next morning, during a second attempt at questioning and prior to counsel being made available to the defendant, the officers obtained incriminating information from the defendant. *Id.* at 478–79. This is further supported by an another Supreme Court decision in which the court observed that once an individual, through his request for counsel, "express[es] his own view that he is not competent to deal with the authorities without legal advice, a later decision at the authorities' insistence to make a statement without counsel's presence may properly be viewed with skepticism." Arizona v. Roberson, 486 U.S. 675, 681 (1988) (quoting Michigan v. Mosley, 423 U.S. 96, 110 n.2 (1975) (White, J., concurring)).

67. New York v. Quarles, 467 U.S. 649, 655 (1984). The Court found that in the interest of public safety, officers should feel free to question suspects concerning potentially dangerous situations and that this *Miranda* exception is available independent of "the motivation of the individual officers involved." *Id.* at 656. The Court explained that the purpose of *Miranda* warnings is to diffuse the coercive pressure of custodial interrogations, which this exception does not undermine. *Id.*

rule protecting the Fifth Amendment's privilege against self-incrimination."[68] In this case, an officer asked the defendant about the location of a gun prior to informing him of his *Miranda* rights.[69] The Court found that locating a potentially deadly weapon in a public location, rather than informing the suspect of his *Miranda* rights, was in the public's best interest.[70] Also, in *Illinois v. Perkins*, the Court held that incriminating statements made by a suspect to his cellmate, who was an undercover officer, were admissible regardless of the fact that the suspect was never informed of his *Miranda* rights prior to any questions the undercover officer asked.[71] In essence, these decisions restricted the use of *Miranda*'s protections as the Court refused to consider expanding *Miranda* past its "underlying" concerns.[72]

Following this period of limitation, *Miranda* began to experience a phase of expansion, with several cases holding in favor of the additional *Edwards* protections.[73] In *McNeil v. Wisconsin*, the Supreme Court stated that *Edwards* was established as a "second layer of prophylaxis for the *Miranda* right to counsel."[74] In *Arizona v. Roberson*, the Court found that *Edwards*, in application, was essentially a continuation of the fundamental purposes of *Miranda* in its protections against the "'inherently compelling pressures' of custodial interrogation and to 'permit a full opportunity to exercise the privilege against self-incrimination.'"[75] Similarly, the Court found that *Edwards* offered the "'clear and unequivocal' guidelines to the law enforcement profession" that *Miranda* equally sought to provide.[76]

With this pro-*Miranda* shift in focus, the Supreme Court further clarified and expanded *Miranda* protections. In *Minnick v. Mississippi*, the

---

68. *Id.* at 657.

69. *Id.* at 652.

70. *Id.* at 657.

71. Illinois v. Perkins, 496 U.S. 292, 300 (1990). The Court, in looking to the fundamental purpose of *Miranda* as avoiding coercive police interrogations, determined that the suspect could never have felt coerced, as he did not know his cellmate's real identity. *Id.* at 296, 298.

72. *Id.* at 296–97; *see also Quarles*, 467 U.S. at 656.

73. *See* McNeil v. Wisconsin, 501 U.S. 171, 176–77 (1991) (summarizing cases that expanded *Edwards*); Arizona v. Roberson, 486 U.S. 675, 682 (1988).

74. *McNeil*, 501 U.S. at 176. In *McNeil*, the Court refused to imply *Miranda* protections from defendants' assertion of their Sixth Amendment right; thus, the defendant's invocation of his Sixth Amendment right did not provide *Edwards*-like protections against his later incriminating statements. *Id.* at 180–81.

75. *Roberson*, 486 U.S. at 681 (quoting Miranda v. Arizona, 384 U.S. 436, 467 (1966)).

76. *Id.* at 682 (quoting *Miranda*, 384 U.S. at 467–68). In *Roberson*, the Court refused to recognize an exception to *Edwards* to question those who have previously asserted a right to counsel (and not yet received it) on unrelated charges. The Court found this proposed extension to be contrary to the goals of explicitly defined police procedures set forth by *Miranda* and *Edwards*. *Id.* at 686.

Court found that *Edwards* protections apply from the time they are initially invoked through all subsequent interrogations.[77] In this case, the defendant asserted his right to counsel in his initial interview.[78] At a subsequent interview, without the presence of counsel, the defendant was informed that he "would 'have to talk'"; he then offered incriminating information.[79] The Court found that if *Edwards* protections were to cease after the initial interrogation, the protection against coercive police procedures would be ineffective, as the coercive atmosphere would reappear upon the second interrogation where counsel may or may not be present.[80] Further, the Court would be left with case-by-case determinations of when, and if, *Edwards* protections applied at the time of questioning, thus eliminating the bright-line ease it sought to provide.[81]

Further support came in *Roberson*, where the Court found that *Edwards* protections span separate investigations, regardless of whether different officers interrogate at different times.[82] The Court held that *Edwards* required the interrogation to cease if counsel is requested, while also expanding its holding to cover questioning about completely separate charges and offenses.[83] In *Roberson*, the defendant, arrested on burglary charges, invoked his right to counsel upon questioning and was subsequently detained.[84] Three days later, a police officer arrived to question the defendant on a different burglary charge.[85] The Court found that the defendant's incriminating statements made in his second interrogation were inadmissible as counsel was not provided to the defendant prior to his second interview.[86] The Court reasoned that, regardless of the nature of the two interrogations, the defendant should have the opportunity to "determine how to deal with the separate investigations with counsel's advice."[87] Cases such as *Edwards*, *Minnick*, and *Roberson* effectively expanded and clarified protections set forth in *Miranda*.[88]

In 2000, the Supreme Court upheld the constitutional basis for *Miranda* and reaffirmed its place in the American judicial process in *Dickerson v. United States*.[89] In *Dickerson*, the Court held that "*Miranda* announced a

77. Minnick v. Mississippi, 498 U.S. 146, 155–56 (1990).
78. *Id.* at 148–49.
79. *Id.* at 149.
80. *Id.* at 154.
81. *Id.* at 154–55.
82. Arizona v. Roberson, 486 U.S. 675, 687–88 (1988).
83. *Id.* at 687.
84. *Id.* at 678.
85. *Id.*
86. *Id.* at 687–88.
87. *Id.* at 687.
88. *See generally* Minnick v. Mississippi, 498 U.S. 146 (1990); *Roberson*, 486 U.S. 675; Edwards v. Arizona, 451 U.S. 477 (1981); Miranda v. Arizona, 384 U.S. 436 (1966).
89. Dickerson v. United States, 530 U.S. 428, 438 (2000).

constitutional rule that Congress may not supersede legislatively,"[90] finding that Congress's confessional "voluntariness" standard of admissibility (found in 18 U.S.C. § 3501) was unconstitutional under *Miranda*.[91] This holding, written by the historically anti-*Miranda* Chief Justice Rehnquist, came as a shock that has intrigued scholars.[92] The *Dickerson* decision managed to preserve "three-and-a-half decades of *Miranda* jurisprudence [that] would have been wiped out."[93]

The *Miranda* rollercoaster continued, however, in the recent cases *Chavez v. Martinez*[94] and *United States v. Patane*.[95] In *Chavez*, the Court reverted to the "prophylactic" nature of *Miranda* and found that a failure to read a suspect his *Miranda* rights did not violate his constitutional rights; instead, it found that the actual *use* of any incriminating information obtained from the subsequent interrogation at trial would violate his Fifth Amendment rights.[96] In her dissent, Justice O'Connor stated that "[t]he Clause must provide more than mere assurance that a compelled statement will not be introduced against its declarant in a criminal trial. Otherwise, there will be too little protection against the compulsion the Clause prohibits."[97] Justice O'Connor disagreed with the majority's conclusion that improper police behavior is permissible if the information obtained from it is not used against the defendant in court.[98]

In *Patane*, the Supreme Court addressed the issue of whether "a failure to give a suspect the warnings prescribed by *Miranda v. Arizona* . . . requires suppression of the physical fruits of the suspect's unwarned-but-voluntary statements."[99] In its analysis, the Court expressly criticized its holding in *Dickerson*, stating that regardless of the "constitutional rule" it lays out, it "does not lessen the need to maintain the closest possible fit between the Self-Incrimination Clause and any judge-made rule designed to protect it."[100] The Court went on to further support its decision in *Chavez* by

---

90. *Id.* at 444.

91. *Id.* at 442–43. 18 U.S.C. § 3501 offered a "voluntariness standard" similar to that of the "totality of the circumstances" test utilized prior to *Miranda*'s bright-line mandate. *See* 18 U.S.C. § 3501 (2006).

92. Kamisar, *supra* note 31, at 200. Kamisar synthesizes various rationales scholars have advanced to account for Rehnquist's position. He notes that Rehnquist's decision may have been: (1) a means to protect its decisions from Congressional imposition; (2) a means to prevent a new era of coercive police procedures; (3) a means to maintain the status quo; or (4) a means to avoid confusion with a reversion to the old "voluntariness standard." *Id.* at 200–01.

93. *Id.* at 201.

94. Chavez v. Martinez, 538 U.S. 760 (2003).

95. United States v. Patane, 542 U.S. 630 (2004).

96. *Chavez*, 538 U.S. at 772–73.

97. *Id.* at 791 (O'Connor, J., dissenting).

98. *Id.* at 795–96.

99. *Patane*, 542 U.S. 633–34.

100. *Id.* at 643.

reaffirming that any information or evidence obtained in violation of *Miranda* would simply not be used in trial.[101]

In his analysis of the recent holdings by the Supreme Court, Professor Kamisar noted, "As *Dickerson* demonstrates, a majority of the Court is unwilling to overrule *Miranda* (or to let Congress do so). As *Patane* makes plain, however, a majority is also unwilling to take *Miranda* seriously."[102]

III. A FOURTEEN-DAY BREAK IN *MIRANDA* CUSTODY TERMINATES *EDWARDS* PROTECTIONS

In *Maryland v. Shatzer*, the Supreme Court, in a 9–0 decision held: (1) once a suspect has experienced a break in custody for longer than fourteen days, protections offered by *Miranda* and *Edwards*, if previously invoked, no longer apply,[103] and (2) "lawful imprisonment imposed upon conviction of a crime does not create the coercive pressures identified in *Miranda*" and therefore the release of a suspect "back into the general prison population" does constitute a break in custody for *Miranda* purposes.[104]

A. Edwards Protections Cease Following a Fourteen-Day Break In Custody Allowing Suspects to Be Reinterrogated

In the majority opinion, written by Justice Scalia, the Court first chose to address the issue of whether a break in custody ends the protections of *Edwards* and, in turn, the protections of *Miranda*.[105] The Court began its analysis by explaining the foundation laid by *Miranda*.[106] Scalia set forth *Miranda*'s most important provisions in noting its effort to "protect a suspect's Fifth Amendment right from the 'inherently compelling pressures' of custodial interrogation."[107] Scalia outlined the elements of the *Miranda* ruling, noting that the suspect has the right to remain silent and the right to have an attorney present–both of which are fully waivable.[108]

Next, the Court noted that this standard of waiver tightened with its holding in *Edwards*, where the Court found that once a suspect invokes his right to counsel, all inquiries must cease "'until counsel has been made available to him, unless the accused himself initiates further

---

101. *Id.*

102. Kamisar, *supra* note 31, at 203.

103. Maryland v. Shatzer, 130 S. Ct. 1213, 1223 (2010).

104. *Id.* at 1224.

105. *Id.* at 1220.

106. *Id.* at 1219.

107. *Id.* (quoting Miranda v. Arizona, 384 U.S. 436, 467 (1966)).

108. *Id.* (citing *Miranda*, 384 U.S. at 444, 475). The Court then notes that a "valid waiver" is attainable through a "knowing, intelligent, and voluntary" standard set forth in Johnson v. Zerbst, 304 U.S. 458 (1938). *Id.* (citing *Miranda*, 384 U.S. at 475).

communication, exchanges, or conversations with the police.'"[109] Scalia went on to explain that the purpose of this is to ensure *Miranda*'s main objective of avoiding coercive police procedures as "subsequent requests for interrogation pose a significantly greater risk of coercion."[110]

After laying this foundation, Scalia analyzed whether a break in custody ends the protections set forth by *Miranda* and *Edwards*, noting that "[l]ower courts have uniformly held that a break in custody ends the *Edwards* presumption, . . . but we have previously addressed the issue only in dicta."[111] Contrary to the *Dickerson* decision, Scalia cited to earlier holdings that noted, "Because *Edwards* is 'our rule, not a constitutional command,' 'it is our obligation to justify its expansion.'"[112] In efforts to rationalize *Edwards*'s expansion to include the break in custody provision, Scalia employed a cost-benefit analysis of prior holdings that employ *Edwards*'s rationale.[113] Scalia, in utilizing *Edwards*'s successor cases, found that "the benefits of the rule are measured by the number of coerced confessions it suppresses that otherwise would have been admitted."[114] Scalia recognized, however, that extending *Edwards* to include a break in custody provision may inadvertently "deter[] law enforcement officers from even trying to obtain" voluntary confessions and further, some truly voluntary confessions will be excluded simply due to their custodial nature.[115]

In determining whether *Edwards*'s protections should include a break in custody exception, Scalia was careful to note that, although "[l]ower courts ha[d] uniformly held that a break in custody ends the *Edwards* presumption,"[116] the Court had "addressed the issue only in dicta."[117] In support of the break in custody provision, Scalia noted that when a suspect is released back "to his normal life for some time before the later attempted interrogation, there is little reason to think that his change of heart regarding interrogation without counsel has been coerced."[118]

---

109. *Id.* (quoting Edwards v. Arizona, 451 U.S. 477, 484–85 (1981)).

110. *Id.* at 1220.

111. *Id.* (citations omitted). The Court referred to its statement in *McNeil* where it stated, "If the police do subsequently initiate an encounter in the absence of counsel (assuming there has been no break in custody), the suspect's statements are presumed involuntary and therefore inadmissible as substantive evidence at trial," McNeil v. Wisconsin, 501 U.S. 171, 177 (1991).

112. *Id.* (quoting Arizona v. Roberson, 486 U.S. 675, 688 (1988) (Kennedy, J., dissenting)).

113. *Id.* at 1220–22.

114. *Id.* at 1220 (citing Montejo v. Louisiana, 129 S. Ct. 2079, 2089 (2009)).

115. *Id.* at 1222.

116. *Id.* at 1220; *see* People v. Storm, 52 P.3d 52, 60–61 n.6 (Cal. 2002) (citing other state and federal cases that have reached the same result).

117. *Shatzer*, 130 S. Ct. at 1220 (citing McNeil v. Wisconsin, 501 U.S. 171, 177 (1991)).

118. *Id.* at 1221.

After his analysis, Scalia concluded that the best means to resolve this dilemma was somehow to limit these protections; otherwise, "every *Edwards* prohibition of custodial interrogation of a particular suspect would be eternal."[119] He noted that, without temporal restriction, the protections could last indefinitely and could apply to any subsequent interrogation, regardless of whether it "pertains to a different crime" or "is conducted by a different law enforcement authority."[120] He went on to observe, "In a country that harbors a large number of repeat offenders, this consequence is disastrous."[121] The Court held:

> The protections offered by *Miranda*, which we have deemed sufficient to ensure that the police respect the suspect's desire to have an attorney present the first time police interrogate him, adequately ensure that result when a suspect who initially requested counsel is reinterrogated after a break in custody that is of sufficient duration to dissipate its coercive effects.[122]

The Court then had to determine what length of time is a "sufficient duration to dissipate [any] coercive effects."[123]

Ultimately, the Court concluded that a time limit must be placed on the protections afforded by *Edwards*, as allowing these protections to last indefinitely and indiscriminately is neither feasible nor does it uphold the underlying concerns of *Miranda* to end the protections upon immediate release.[124] The Court concluded that a specific time limit was necessary to ensure the correct result.[125] Ultimately, the Court deemed the appropriate time period to be fourteen days, reasoning that it "provides plenty of time for the suspect to get reacclimated to his normal life, to consult with friends and counsel, and to shake off any residual coercive effects of his prior custody."[126] The Court contended that this bright-line rule would save time and judicial resources in that, if a fourteen-day break in custody occurs, no "fact-intensive inquiry" is necessary to determine when and if the suspect previously exercised his right to counsel as the protections no longer apply.[127]

---

119. *Id.* at 1222.
120. *Id.*
121. *Id.*
122. *Id.*
123. *Id.*
124. *Id.* at 1223. Instead, "the police will release the suspect briefly (to end the *Edwards* presumption) and then promptly bring him back into custody for reinterrogation." *Id.*
125. *Id.*
126. *Id.*
127. *Id.* at 1224.

In his concurrence, Justice Thomas agreed with the Court's overall judgment but disagreed with the Court's fourteen-day period.[128] Thomas criticized the overall extension of *Edwards* outside the specific conditions of the *Edwards* case, however, he reasoned that, even if the holding of *Edwards* does extend to other cases, "the Court's 14-day rule fails to satisfy the criteria our precedents establish for the judicial creation of such a safeguard."[129] Thomas contended that precedential law already provided for this situation.[130] In his analysis, Thomas attempted to rationalize the fourteen-day rule through his own cost-benefit analysis, contending that the fourteen-day rule lacks a concrete basis for its specificity in duration.[131]

Justice Stevens also concurred with the overall judgment but disagreed with the rule's bright-line application.[132] Similarly to Justice Thomas, Stevens noted that this rule "disregards much of the analysis upon which *Edwards* and subsequent decisions were based."[133] Stevens concluded that the fourteen-day rule was arbitrary and that the majority's assumption—that once a suspect is released from custody, exposure to others outside of prison will aid the suspect through effective advice—is "overconfident and . . . questionably relevant."[134]

Scalia responded to Stevens's comments by noting that, "[i]nstead of terminating *Edwards* protection when the custodial pressures that were the basis for the protection dissipate, the concurrence would terminate it when the suspect would no longer 'feel that he has been denied the counsel he has clearly requested.'"[135] Scalia noted that, as the provision of counsel is not the central point of *Edwards*,[136] Stevens's rationale that counsel is the breaking point for *Edwards* is flawed.[137] Further, Scalia commented on Stevens's analysis of the fourteen-day break period by saying, "Failure to say where the line falls short of 2 1/2 years, and leaving that for future case-by-case determination, is certainly less helpful, but not at all less arbitrary."[138]

---

128. *Id.* at 1227 (Thomas, J., concurring).

129. *Id.*

130. *Id.*; *see* Johnson v. Zerbst, 304 U.S. 458, 464 (1938) (finding that "[t]he determination of whether there has been an intelligent waiver of the right to counsel must depend, in each case, upon the particular facts and circumstances surrounding that case, including the background, experience, and conduct of the accused.").

131. *Shatzer*, 130 S. Ct. at 1228 (Thomas, J., concurring).

132. *Id.* (Stevens, J., concurring).

133. *Id.* at 1230 (Stevens, J., concurring). Stevens notes that prior cases in the "*Miranda* line of cases" already sufficiently provide protection. *Id.* (citing Michigan v. Harvey, 494 U.S. 344 (1990); Minnick v. Mississippi, 498 U.S. 146 (1990); Arizona v. Roberson, 486 U.S. 675 (1990)).

134. *Id.* at 1231–32 (Stevens, J., concurring).

135. *Id.* at 1226 (majority opinion) (quoting *id.* at 1234 (Stevens, J., concurring)).

136. *Id.*

137. *Id.*

138. *Id.*

B. Release Back Into the General Prison Population Constitutes a Break in Miranda Custody

In *Maryland v. Shatzer*, the Court took the case a step further in recognizing a second important issue: does releasing a suspect back into the general prison population, where he is serving an unrelated sentence, constitute a break in *Miranda* custody?[139] The Court noted that, "[t]o determine whether a suspect was in *Miranda* custody we have asked whether 'there is a formal arrest or restraint on freedom of movement of the degree associated with a formal arrest.'"[140] The Court further noted that, because "all forms of incarceration" clearly satisfy this test, it is only a "necessary" but not "sufficient condition for *Miranda* custody."[141] In reverting back to the original purposes of *Miranda*, the Court held, "[W]e think lawful imprisonment imposed upon conviction of a crime does not create the coercive pressures identified in *Miranda*."[142]

Scalia rationalized his decision by noting that prison is where convicted criminals live, interact, and communicate with others.[143] He contended that the coercive pressures of continued, isolated interrogation do not exist when one is released back into their normal prison environment surrounded by familiar "inmates, guards, and workers."[144] Regarding Shatzer's case, Scalia noted that "[t]he 'inherently compelling pressures' of custodial interrogation ended when he returned to his normal life" in prison.[145]

In Stevens's concurrence, he found the Court's description of a prisoner's life unrealistic, stating that "[p]risoners are uniquely vulnerable to the officials who control every aspect of their lives; prison guards may not look kindly upon a prisoner who refuses to cooperate with police."[146] Stevens then noted that police may take advantage of a prisoner's situation in that they neither have to promptly question nor "initiate formal legal proceedings" with the suspect who is already detained.[147] Stevens concluded that, although he disagreed with the specific time limit and

---

139. *Id.* at 1224.

140. *Id.* (quoting New York v. Quarles, 467 U.S. 649, 655 (1984)).

141. *Shatzer*, 130 S. Ct. at 1224.

142. *Id.* at 1225.

143. *Id.*

144. *Id.* The Court further noted that inmates at facilities like Shatzer's can usually, "visit the library each week," "have regular exercise and recreation periods," "participate in basic adult education and occupation training," "send and receive mail," and "receive visitors twice a week." *Id.* at 1225 (citations omitted).

145. *Id.* at 1224.

146. *Id.* at 1233 (Stevens, J., concurring). Stevens commented that prisoners may not be as comfortable as the majority believes. He noted that "cooperation frequently is relevant to whether the prisoner can obtain parole." *Id.* Stevens further contended that a suspect already imprisoned may feel controlled by those questioning him and may be unable to differentiate between the two instances, his imprisonment and his questioning. *Id.*

147. *Id.* at 1233 n.13.

believed that a release back into a prison was not a break in custody, he concurred in the overall judgment.[148] He reasoned that, in Shatzer's specific situation, "this period of time[149] was sufficient" in that "the second interrogation [was] no more coercive than the first."[150]

IV. CONSEQUENCES OF MARYLAND V. SHATZER

In *Maryland v. Shatzer*, the Court's analysis and holding continued the forty-year trend of *Miranda*'s expansion and limitation. It is noteworthy that, while *Shatzer* expanded the provisions, it followed two cases in which the Court declined to extend *Miranda*.[151] This continues the Court's fickle treatment of these safeguards.

It is first interesting to note that *Shatzer* decision did not squarely address the legitimacy of the holding in *Dickerson* regarding the constitutional power of the *Miranda* safeguards. In *Shatzer*, the Court continued to relate that "*Edwards* is 'our rule, not a constitutional command,' 'it is our obligation to justify its expansion.'"[152] This, like other cases after *Dickerson*, seemed to directly contradict the Court's holding in *Dickerson*.[153]

As in previous *Miranda* cases, the Court was correct in remaining conscious of *Miranda*'s underlying concerns of coercive police behavior. In its holding regarding a break in custody exception, the Court's main goal was to "dissipate [the] coercive effects" of custodial interrogation.[154] Second, in finding that a prisoner's return to the general prison population constituted break in custody, the Court found that "[t]he 'inherently compelling pressures' of custodial interrogation ended when [Shatzer] returned to his normal life."[155] As the Court employed this rationale in prior *Miranda* extensions, it properly continued its course in *Shatzer*.

The Court's analysis of the "break in custody" issue remained well supported and perhaps overdue. As the Court noted, lower courts had previously implemented this idea, while the Court has remained relatively silent.[156] The Court's analysis seemed to indicate that it had already made

148. *Id.* at 1234.

149. "2 1/2 years." *Id.*

150. *Id.*

151. *See generally* United States v. Patane, 542 U.S. 630 (2004); Chavez v. Martinez, 538 U.S. 760 (2003) (declining to extend *Miranda*'s protections).

152. *Shatzer*, 130 S. Ct. at 1220 (quoting Arizona v. Roberson, 486 U.S. 675, 688 (1985) (Kennedy, J., dissenting)).

153. Dickerson v. United States, 530 U.S. 428, 428 (2000). *Dickerson* held that *Miranda*'s holding was a constitutional provision and could not be superseded by acts of Congress via legislation that was contradictory to its holding. *Id.* at 444.

154. *Shatzer*, 130 S. Ct. at 1222.

155. *Id.* at 1225.

156. *Id.* at 1220.

this decision and that *Shatzer* was a means by which to officially declare it.[157] While the Court's cost-benefit analysis of the break in custody issue remains persuasive, its overall balance of the factors may be closer than the Court alluded. The majority held that employing this break in custody rule would benefit the overall custodial situation by "'conserv[ing] judicial resources,'"[158] maintaining the individual's "'choice to communicate with police only through counsel,'"[159] and preventing police "'badgering.'"[160] The Court contended that these benefits outweigh the drawbacks, which the Court deemed to be fewer voluntary confessions and the deterrence of law enforcement officers from seeking confessions.[161] However, the benefits of this bright-line rule may not be as definitive as the Court concluded.

In regard to preserving resources, while the courts may save time and money, the police force may find this bright-line rule a mechanism by which to simply wait two weeks to reinterrogate a suspect. In his concurrence, Justice Thomas pointed out that "the Court's time-based rule . . . disregards the compulsion caused by a second (or third, or fourth) interrogation of an indigent suspect who was told that if he requests a lawyer, one will be provided for him."[162] This point demonstrates that perhaps the rule will not prevent police badgering, and while it may save judicial resources, it may nevertheless require more public resources when police choose to reinterrogate suspects in two-week intervals.

While the Court maintained its ability to impose "precise time limits governing police action,"[163] its arbitrary implementation of fourteen days is troubling. As Justice Stevens's concurrence noted, the Court's determination lacked any kind of statistical support or evidence in finding fourteen days adequate to alleviate the worry of coercive police procedure.[164] The Court found that "[c]onfessions obtained after a 2-week break in custody and a waiver of *Miranda* rights are most unlikely to be compelled, and hence are unreasonably excluded."[165] It is well argued that a limitation must be found for the *Edwards* protection, as it cannot last indefinitely.[166] Further, a case-by-case determination would leave the Court vulnerable to inconsistent results.[167] The Court's position might have been

157. *Id.* (citing McNeil v. Wisconsin, 501 U.S. 171, 177 (1991)).

158. *Id.* (quoting Minnick v. Mississippi, 498 U.S. 146, 151 (1990)).

159. *Id.* (quoting Patterson v. Illinois, 487 U.S. 285, 291 (1988)).

160. *Id.* (quoting Michigan v. Harvey, 494 U.S. 344, 350 (1990)).

161. *Id.* at 1222.

162. *Id.* at 1229 (Thomas, J., concurring).

163. *Id.* at 1223; *see* County of Riverside v. McLaughlin, 500 U.S. 44, 57, 58–59 (1991) (where the Court imposed a forty-eight-hour time limit on a certain police procedure).

164. *Shatzer*, 130 S. Ct. at 1231 n.7 (Stevens, J., concurring).

165. *Id.* at 1223.

166. *Id.* at 1222.

167. *Id.*

stronger had it offered better evidence that fourteen days is the optimum threshold, perhaps a statistical analysis or simply a more thorough explanation. However, the Court offers only broad, unsupported generalizations.[168]

The Court's analysis of whether a release back into a prison population constitutes a break in custody is better supported with evidence and explanation. The Court's explanation employed factual prison information of a prisoner's daily life and how prisoners actually find a "home" in prison.[169] However, the Court was again very general in its analysis. Stevens's disagreement on this issue went past the mere physical privileges in prison and looked to the psychological effects that the Court should have considered.[170] Yet again, the Court overgeneralized the situation, indicating that prisoners receive many of the same freedoms as those who are not in prison,[171] while completely ignoring important differences that may affect the situation. The Court failed to recognize that not every prisoner's situation is the same and that, although prison may be the prisoner's "normal life,"[172] it is still run by others who maintain control and power over the prisoner.[173] Instead, the Court could have set forth specific factors for determining whether an individual experiences a break in *Miranda* custody. With such factors, the Court could analyze the extent to which the suspect may employ free will and decision-making uncontaminated by coercive police power. Generally, the break in custody exception should apply to prisoners because, as the Court noted, certain situations do exist in which prisoners enjoy many of the same freedoms as non-prisoners.[174] However, for the Court to generalize that all prisoners enjoy the freedoms mentioned is inconsiderate. With the current opinion set forth by the Court, prisoners who find themselves in an isolated situation are grouped with those who enjoy access to outside materials, people, and ideas.

The Court's decision provided this area of law with a bright-line analysis and application, yet its broad generalizations and its imposition of arbitrary time limits remain disconcerting. While a break in custody seems a logical punctuation to one's *Edwards* protections, as the Court noted, its

---

168. *See id.* at 1228 (Thomas, J., concurring); *see also id.* at 1231 n.7 (Stevens, J., concurring).

169. *Id.* at 1224–25.

170. *Id.* at 1232–33 (Stevens, J., concurring). Stevens noted that "for a person whose every move is controlled by the State, it is likely that 'his sense of dependence on, and trust in counsel as the guardian of his interests in dealing with government officials intensified.'" *Id.* (quoting United States v. Green, 592 A.2d 985, 989 (D.C. 1991)).

171. *Id.* at 1224–25. The Court focuses on the many privileges given to some prisoners, such as: "visit[ing] the library each week," "hav[ing] regular exercise and recreation periods," "participat[ing] in basic adult education and occupation training," "send[ing] and receiv[ing] mail," and "receiv[ing] visitors twice a week." *Id.* at 1225.

172. *Id.* at 1225.

173. *Id.* at 1233 (Stevens, J., concurring).

174. *Id.* at 1225.

application requires structure.[175] The Court found that the fourteen-day time limit best achieved this structure.[176] Further, the Court found that a break in custody benefits individuals by allowing them to return to their normal life, which the Court finds includes sentenced prisoners.[177]

V. CONCLUSION

Maryland v. Shatzer continues the epic forty-year debate concerning the protections offered by *Miranda* and how much the Court should extend or limit its application. This case demonstrates both the complexity of years of constitutional precedent and how the Court must try to tie it all together in formulating a decision. The Court's holding guides law enforcement, lawyers, and judges in determining whether a suspect's Fifth Amendment rights were and are protected. However, the rule's strict application leaves little room for interpretation and consideration of differing circumstances. The holding will encourage law enforcement to follow the bright-line rules, yet it does little to protect against new forms of coercive police behavior. Further, the holding will ensure that prisoners, although incarcerated, will not enjoy rights that law-abiding citizens do not. However, it may remain unfair to those prisoners who do not enjoy the freedoms of this "normal life" argued by the Court. We can only wait and see what effect future application and interpretation of these rules will have on the overall protections offered by *Miranda*.

CAITLYN LUEDTKE ELAM

175. *Id.* at 1222.
176. *Id.* at 1223.
177. *Id.* at 1225.

CONSTITUTIONAL LAW—SIXTH AMENDMENT RIGHT TO COUNSEL— FORFEITURE OF THE RIGHT TO COUNSEL IN TENNESSEE

State v. Holmes, 302 S.W.3D 831 (TENN. 2010).

I. INTRODUCTION

In 2003, the Criminal Court for Shelby County appointed an assistant public defender ("Counsel") to represent Tommy Holmes ("the Defendant") following his indictment for attempted murder, aggravated rape, intimidation of a witness, and aggravated assault.[1] As a result of two incidents during meetings with the Defendant, Counsel requested the trial court's permission to withdraw from the case at an "impromptu hearing" about one week before the Defendant's trial.[2] Counsel explained that during a meeting with the Defendant earlier in the day, the Defendant poked Counsel in the face with his finger, knocking Counsel's glasses off.[3] In addition, the Defendant threatened physical violence at their previous meeting.[4] Upon hearing these accusations, the trial court allowed Counsel to withdraw without hearing further testimony[5] and concluded a hearing on the matter was not necessary.[6] The court also determined on its own that the Defendant waived his right to court-appointed counsel[7] because of his conduct towards Counsel.[8] As a result, he would have to represent himself at trial unless he or his family hired an attorney.[9] Accordingly, the court denied the Defendant's ensuing written request for counsel.[10]

---

1. State v. Holmes, 302 S.W.3d 831, 834 (Tenn. 2010).
2. *Id.*
3. *Id.*
4. *Id.*
5. *Id.* The defendant was present at this "impromptu hearing" but did not testify. *Id.*
6. State v. Holmes, No. W2006-00236-CCA-R3-CD, 2007 WL 1651876, at *3 (Tenn. Crim. App. June 7, 2007). The trial court judge stated: "I don't find there to be a need to have a formal hearing in this regard. [Appointed counsel] is a lawyer that's practiced in this court for many years. I know him to be a man of impeccable integrity, and if he says that happened, as an officer of this court, then I accept his statement that that is what happened." *Id.* (alteration in original).
7. *Holmes*, 302 S.W.3d at 834. The Sixth Amendment of the United States Constitution and Article 1, section 9 of the Tennessee Constitution guarantee an indigent criminal defendant the right to the assistance of court appointed counsel. *See* Gideon v. Wainwright, 372 U.S. 335, 344 (1963); State v. Northington, 667 S.W.2d 57, 60 (Tenn. 1984).
8. *Holmes*, 2007 WL 1651876, at *4. The trial court judge declared: "If a defendant chooses to take action such as [the defendant] took today, he does so at his [own] peril and the result is he is going to be representing himself . . . ." *Id.* (alteration in original).
9. *Id.* ("So, from this point forward, unless you're able to hire a lawyer . . . you're

At a hearing on pretrial motions almost two months later, the Defendant denounced Counsel's previous allegations as false and once more expressed his need for an attorney.[11] The Defendant declared that he and Counsel were not alone at the time that he allegedly assaulted Counsel and that others could rebut Counsel's allegations.[12] Choosing to rely solely on Counsel's accusations,[13] the trial court again denied the Defendant's request for an attorney and refused to hold a hearing on the alleged incident.[14] Consequently, the Defendant represented himself[15] at his jury trial on the aggravated rape charge.[16] The jury convicted the Defendant of aggravated rape,[17] and the Defendant received a sentence of twenty-four years in the Department of Correction.[18]

Subsequently, the Defendant appealed his conviction with the assistance of counsel.[19] On appeal, the Tennessee Court of Criminal Appeals determined that the trial court erred by determining that the Defendant waived his right to counsel without holding an evidentiary hearing on the alleged incident.[20] Because the trial court did not give the Defendant an opportunity to testify about what occurred during his last meeting with Counsel, the appellate court remanded the matter to the trial court for a hearing to determine whether the Defendant waived his right to counsel.[21]

---

going to be representing yourself.").

10. *Id.*

11. *Id.*

12. *Id.* At the hearing on pretrial motions the defendant claimed that "a 'sheriff' and 'plenty [of] inmates' were present." *Id.* (alteration in original).

13. *Id.* The trial judge asserted that a hearing was not necessary because "when it comes down to the final analysis, I believe [appointed counsel] is telling me the truth." *Id.* (alteration in original).

14. *Id.*

15. *Id.* Although the Defendant was forced to represent himself, "the trial court subsequently appointed advisory or 'elbow' counsel to assist him in conducting his defense, ordering that the advising attorney would 'be available to talk to [the defendant] and confer with [him] during recesses, but [could] not sit at counsel table with [him] during the trial.'" *Id.* (alteration in original). A defendant may confer with elbow counsel "'for guidance and advice, but otherwise handles the defense of the case on his or her own'" because elbow counsel "'functions in a purely advisory role, without actively participating in the trial.'" State v. Holmes, 302 S.W.3d 831, 835 n.3 (Tenn. 2010) (quoting State v. Small, 988 S.W.2d 671, 672 n.1 (Tenn. 1999)).

16. State v. Holmes, 302 S.W.3d 831, 834 (Tenn. 2010).

17. *Id.* The defendant's trial began in September 2005. *Id.* at n.2.

18. *Holmes*, 2007 WL 1651876, at *1. The defendant also represented himself at his sentencing hearing. *Id.* at *4.

19. *Id.* The defendant's elbow counsel from his aggravated rape trial was allowed to directly represent him on appeal. *Id.*

20. *Holmes*, 302 S.W.3d at 834.

21. *Holmes*, 2007 WL 1651876, at *6. The court explained that "in light of the defendant's assertions that there were witnesses to the altercation, we must remand for

At the hearing on remand,[22] Counsel testified that at the conclusion of his penultimate meeting with the Defendant, the Defendant said, "I know how to get rid of you."[23] Counsel stated that although the Defendant's statement "sounded threatening," Counsel chose not to report the Defendant's statement or request that a guard handcuff the Defendant at future meetings.[24] Furthermore, Counsel testified that the Defendant "pushed his finger" at Counsel knocking his glasses askew during their last meeting.[25] Counsel stated that although the Defendant made contact with his eyeglasses rather than his face, the Defendant's contact was painful and broke the glasses at the temple.[26] Counsel admitted to being frightened by the Defendant's actions,[27] especially in light of recent acts of violence against attorneys in Shelby County.[28] A bailiff who spoke with Counsel immediately after the incident testified that Counsel stated that the Defendant had "slapped" him, but that he did not wish to press charges against him.[29] The Defendant was present at the hearing, but chose not to testify.[30] Left with no evidence other than the testimony of Counsel and the bailiff,[31] the judge ruled that the State had "met its burden of demonstrating" that the Defendant waived his right to the assistance of appointed counsel by physically attacking Counsel.[32]

The Defendant appealed his conviction, and a split Tennessee Court of Criminal Appeals affirmed the decision of the trial court,[33] upholding the

---

consideration of this issue." *Id.* On remand, "the State has the burden of establishing that the defendant forfeited his right to counsel." *Id.*

22. *Holmes*, 302 S.W.3d at 834. A different trial judge conducted the hearing because the original trial judge had retired. *Id.*

23. *Id.* at 835.

24. Brief of the State of Tennessee at 6, State v. Holmes, 302 S.W.3d 831 (Tenn. 2010) (No. W2008-00759-SC-R11-CD).

25. *Holmes*, 302 S.W.3d at 835. The defendant assaulted Counsel after Counsel declined to state whether he believed the defendant was innocent. *Id.*

26. State v. Holmes, No. W2008-00759-CCA-R3-CD, 2009 WL 536930, at *2 (Tenn. Crim. App. Mar. 2, 2009).

27. *Holmes*, 302 S.W.3d at 835. Counsel described the Defendant's attack as "astounding" and reported it to the trial court without delay. *Id.*

28. *Id.* at 836. Counsel had previously assisted in representing Tony Carruthers, a convicted murder who threatened and harassed Counsel and another attorney at trial. *Id.* at n.4. Memphis attorney Robert Friedman was infamously killed by a disgruntled client in 2002. *Id.*

29. *Holmes*, 2009 WL 536930, at *2. Counsel did not wish to press charges because he believed the Defendant "was in too much deep water as it is." *Id.*

30. *Holmes*, 302 S.W.3d at 836.

31. Brief of the State of Tennessee, *supra* note 24, at 8.

32. *Holmes*, 2009 WL 536930, at *3. The judge stated that the "finding of fact is that Mr. Holmes punched [Counsel] in the face with his finger and made contact with his glasses on his face." Brief of the State of Tennessee, *supra* note 24, at 9.

33. *Holmes*, 302 S.W.3d at 837.

Defendant's aggravated rape conviction.[34] Specifically, the majority concluded that "the [D]efendant's physical assault against counsel, combined with his earlier verbal threat, qualifies as the sort of extremely serious misconduct sufficient to warrant the trial court's finding that the defendant forfeited his right to the assistance of counsel."[35] The Defendant then appealed to the Tennessee Supreme Court.[36] On appeal, the Court, *held*, reversed and remanded.[37] The Defendant's conduct was not outrageous enough to merit the "extreme sanction" of forfeiture of the right to appointed counsel without prior warning because of the following circumstances: (1) the Defendant's forfeiture deprived him of his right to counsel at trial, (2) the Defendant was not attempting to "obstruct, delay or manipulate the proceedings," (3) the Defendant's attack was not violent and caused no bodily injury, (4) the Defendant's assault consisted of a single attack on his original attorney, and (5) measures other than forfeiture could have been utilized to protect Counsel.[38]

II. CONDUCT SUFFICIENT TO WARRANT FORFEITURE OF COUNSEL

In *State v. Carruthers*, the only other Tennessee Supreme Court case analyzing a defendant's forfeiture of the right to appointed counsel,[39] the defendant both implicitly waived and forfeited the right to counsel after the trial court warned him that forfeiture of counsel would result from continued misconduct.[40] Because the defendant continued to misbehave in a manner that was sufficient for a finding of forfeiture after receiving a warning, the *Carruthers* court ruled that the defendant had both implicitly waived and forfeited his right to counsel.[41] In the instant case, the Defendant received no such warning before his forfeiture of counsel, and accordingly, the Defendant could not have implicitly waived his right to counsel.[42] As a result, the Court faced the question of whether the Defendant's misconduct was reprehensible enough to warrant forfeiture of counsel in the absence of a prior warning from the trial court informing the

34. *Holmes*, 2009 WL 536930, at *5.

35. *Id.* The dissenting judge did not agree that the Defendant's conduct "rose to the level of extremely serious misconduct to warrant forfeiture of his right to counsel," and found that the trial court abused its discretion in ruling that the defendant had waived his right to counsel. *Id.* at *6–7 (Thomas, Jr., J., dissenting).

36. *Holmes*, 302 S.W.3d at 837.

37. *Id.* at 848. The court reversed and remanded because the trial court's finding that the defendant had forfeited his right to trial was an "error [that] compromised the integrity" of the defendant's trial making it "a structural constitutional error." *Id.*

38. *Id.* at 847–48.

39. *Id.* at 839 (referencing State v. Carruthers, 35 S.W.3d 516 (Tenn. 2000)).

40. *Carruthers*, 35 S.W.3d at 549–50.

41. *Id.* at 550.

42. *Holmes*, 302 S.W.3d at 841.

Defendant of the consequences of further misbehavior.[43] In addition, the immediate case also differs from *Carruthers* because the Defendant's extensive misconduct in *Carruthers* included a "significant pattern of verbal threats" and caused "the withdrawal of seven lawyers and deliberate delay of the judicial process."[44] Accordingly, the unique question before the Tennessee Supreme Court in *Holmes* was whether the Defendant's single attack, which caused no bodily injury, and lone ambiguous threat against his first and only appointed attorney constituted conduct extreme enough to justify forfeiture of the fundamental right to counsel.[45]

III. RELINQUISHMENT OF THE RIGHT TO COUNSEL

A. Introduction

The United States and Tennessee Constitutions guarantee that indigent defendants have the assistance of counsel in state criminal prosecutions.[46] Although the right to counsel is a fundamental right,[47] defendants can relinquish their appointed counsel in three ways.[48] First, defendants can knowingly and voluntarily waive their right to counsel and affirmatively choose to represent themselves.[49] Second, defendants can implicitly waive their right to counsel through repeated misconduct, particularly conduct that delays or disrupts proceedings.[50] A defendant who is explicitly warned that further misconduct may result in losing the right to counsel yet chooses to continue misbehaving implicitly and involuntarily waives the right to counsel.[51] Waiver by conduct occurs only when the trial court explicitly warns the defendant of the consequences of continued misconduct, including the dangers of self-representation as a layman.[52] Finally, a defendant can forfeit his right to counsel by engaging in misconduct that is so egregious that the trial court need not warn the defendant before denying the right to counsel.[53] Forfeiture differs from waiver in that waiver entails a

---

43. *Id.* The court recognized that "only the most egregious misbehavior will support a forfeiture of that right without warning and an opportunity to conform his or her conduct to an appropriate standard." *Id.* at 846.

44. *Id.* at 841.

45. *Id.*

46. U.S. CONST. amend. VI; TENN. CONST. art. 1, § 9; Gideon v. Wainwright, 372 U.S. 335, 344 (1963); State v. Northington, 667 S.W.2d 57, 60 (Tenn. 1984).

47. Johnson v. Zerbst, 304 U.S. 458, 462 (1938).

48. Suzanne Diaz, Case Note, State v. Hampton: *Addressing Forfeiture of the Right to Counsel by Egregious Conduct*, 47 ARIZ. L. REV. 837, 838 (2005) (citing State v. Hampton, 92 P.3d 871, 873–74 (Ariz. 2004)).

49. *Id.* at 838–39 (citing Faretta v. California, 422 U.S. 806, 835 (1975)).

50. *Id.* at 839 (citing United States v. Goldberg, 67 F.3d 1092, 1100 (3d Cir. 1995)).

51. *Id.* (citing *Goldberg*, 67 F.3d at 1100).

52. *Id.* (citing *Hampton*, 92 P.3d at 874).

53. *Id.* (citing *Hampton*, 92 P.3d at 874).

knowing and intentional relinquishment of a known right,[54] whereas forfeiture occurs even though a defendant may not have intended to waive the right to counsel or understood the risks of self-representation.[55] While the United States Supreme Court has never declared what circumstances will result in forfeiture of appointed counsel,[56] the Court has ruled that defendants can forfeit, without warning, other trial-related constitutional rights because of misconduct.[57]

B. Forfeiture Analysis in Federal Courts

Although the United States Supreme Court has not yet recognized more than one form of waiver of the Sixth Amendment,[58] federal appellate courts, starting with the Eleventh Circuit in *United States v. McLeod*, have recognized both implicit waiver and forfeiture.[59] In *United States v. McLeod*, the Eleventh Circuit distinguished involuntary forfeiture of counsel from voluntary waiver of counsel and ruled that defendants who maliciously abuse and threaten their appointed counsel can forfeit their right to counsel without prior warning.[60] The defendant in *McLeod* threatened to harm his second appointed attorney and verbally abused him during a telephone conversation.[61] Furthermore, the defendant, who had previously dismissed his first attorney, threatened to sue his second attorney on four separate occasions and tried to coerce the attorney into using unethical defense tactics.[62] The court asserted that although the right to counsel is "cherished" and "fundamental," it "may not be put to service as a means of delaying or trifling with the court."[63] Noting that courts in other situations have ruled that a defendant "may forfeit constitutional rights by virtue of his or her actions,"[64] the court concluded that the defendant similarly forfeited his right to counsel as a result of his pattern of disruptive

---

54. Jennifer Elizabeth Parker, Casebrief, *Constitutional Law*—United States v. Goldberg: *The Third Circuit's Nontraditional Approach to Waiver of the Sixth Amendment Right to Counsel*, 41 VILL. L. REV. 1173, 1209 n.189 (1996) (citing *Goldberg*, 67 F.3d at 1100–1101).

55. Parker, *supra* note 54, at 1204 (citing *Goldberg*, 67 F.3d at 1100).

56. Gilchrist v. O'Keefe, 260 F.3d 87, 94 (2d Cir. 2001). The United States Supreme Court has "failed to recognize more than one category of waiver" and does not differentiate between voluntary waiver of counsel and involuntary forfeiture of counsel. Parker, *supra* note 54, at 1196 (citing *Goldberg*, 67 F.3d at 1100).

57. *See Goldberg*, 67 F.3d at 1100–01 (citing Illinois v. Allen, 397 U.S. 337, 343 (1970)).

58. Parker, *supra* note 54, at 1196 (citing *Goldberg*, 67 F.3d at 1100).

59. *Id.* (citing United States v. McLeod, 53 F.3d 322 (11th Cir. 1995)).

60. *McLeod*, 53 F.3d at 325.

61. *Id.*

62. *Id.*

63. *Id.* (quoting United States v. Fowler, 605 F.2d 181, 183 (5th Cir. 1979)).

64. *Id.*; *see* Illinois v. Allen, 397 U.S. 337, 343 (1970); *Fowler*, 605 F.2d at 183.

and insulting behavior and upheld the trial court's ruling that the defendant forfeited his right to counsel on his motion for a new trial.[65] Significantly, the court affirmed the defendant's forfeiture despite the fact that the trial court did not first warn the defendant that continued misbehavior would result in the loss of appointed counsel.[66] Moreover, the court recognized a critical distinction between forfeiture and waiver of counsel by acknowledging that because the defendant still desired the assistance of counsel, he was involuntarily forfeiting his right to counsel rather than voluntarily waiving it.[67]

Federal courts have also ruled that defendants who violently attack their appointed attorneys without provocation forfeit the right to counsel.[68] In *United States v. Jennings*, a defendant who hit his appointed counsel in the side of the head with such force that the attorney fell to the floor was denied his right to counsel.[69] The defendant also threatened to kill his attorney, the corrections officers, and the prosecutor.[70] The United States District Court for the Middle District of Pennsylvania asserted that because "[n]o court can carry on its business in an atmosphere of violence, fear and intimidation," violence at trial is a threat not only to the safety of attorneys but also to the operation of the judicial system and cannot be tolerated.[71] Therefore, in the court's view, defendants who grossly exploit their right to counsel by attacking their attorney in order to achieve their goal of obtaining a new attorney should lose the right to counsel.[72] Accordingly, the court determined that the defendant's violent behavior was sufficiently "extreme and outrageous" as to warrant forfeiture of counsel.[73] Recognizing that forfeiture of counsel is "an extreme sanction," the court reasoned that

---

65. *McLeod*, 53 F.3d at 326. "Our holding is limited to [the defendant's] forfeiture of his right to counsel at the hearing on the motion for a new trial." *Id.* at n.13.

66. Parker, *supra* note 54, at 1197. The court stated that "[i]n light of [defendant's] behavior, we cannot say that the district judge erred by concluding that [defendant] had forfeited this right to counsel." *McLeod*, 53 F.3d at 326. Although the trial court did not warn the defendant before he forfeited the right to counsel, the trial court did hold a hearing on the attorney's motion to withdraw in which the defendant was invited to testify but declined. *Id.*

67. *McLeod*, 53 F.3d at 326.

68. United States v. Jennings, 855 F. Supp. 1427, 1445 (M.D. Pa. 1994), *aff'd*, 61 F.3d 897 (3d Cir. 1995) (unpublished table decision).

69. *Id.* at 1432, 1445. Six federal marshals were necessary to control the defendant. *Id.* at 1433.

70. *Id.*

71. *Id.*

72. *Id.* at 1444. The court held "that an indigent defendant who, without provocation or justification, physically assaults court-appointed counsel, thereby waives the right to appointed counsel." *Id.* at 1445.

73. *Id.* at 1444. The Third Circuit affirmed the court's ruling in an unpublished decision. United States v. Jennings, 61 F.3d 897 (3d Cir. 1995).

forfeiture of counsel is an appropriate response to egregious, violent misconduct by defendants.[74]

In *United States v. Goldberg*, the Third Circuit first defined "voluntary waiver," "forfeiture," and "implicit waiver" of the right to counsel[75] and differentiated these three separate processes by which an indigent defendant can lose the right to counsel.[76] Declaring that forfeiture of the right to counsel requires a demonstration of "extremely serious misconduct,"[77] the court explained that implicit waiver of counsel "requires that a defendant be warned about the consequences of his conduct, including the risks of proceeding pro se," whereas no warning or disclaimer of the risks is necessary for forfeiture.[78] Because defendants are warned about the consequences of their conduct in the instance of an implicit waiver, the court surmised that the misconduct required for implicit waiver is not as severe as that required for forfeiture.[79] Furthermore, the court indicated that forfeiture of the right to counsel is an extreme sanction that should require correspondingly extreme misconduct.[80] Although the court found that there was no implicit waiver because the trial court did not properly warn the defendant,[81] the court suggested that the defendant's intentional efforts to delay the proceedings would have been sufficient for the trial court to find implicit waiver if the warning had been proper.[82] As for forfeiture, the court declined to determine whether the defendant's death threat was extreme enough to warrant forfeiture because the defendant had not been present at the hearing in which the trial court determined that the defendant had indeed made threatening statements to his attorney.[83] As a result, the State could not use the death threat to demonstrate egregious conduct

---

74. *Jennings*, 63 F.3d at 1444. The court noted that employing punitive measures like fines or additional jail time to deter violence likely would have little effect on an indigent inmate serving a life sentence such as the defendant. *Id.*

75. Parker, *supra* note 54, at 1207–08.

76. United States v. Goldberg, 67 F.3d 1092, 1099–1102 (3d Cir. 1995). The court recognized "an important distinction between the ideas of "waiver" and "forfeiture," and a hybrid of those two concepts, "waiver by conduct." *Id.* at 1099.

77. *Id.* at 1102. The court stated that forfeiture can also occur upon a showing of "extremely dilatory conduct." *Id.* at 1101.

78. *Id.*

79. Parker, *supra* note 54, at 1205 (citing *Goldberg*, 67 F.3d at 1101).

80. *Id.* (citing *Goldberg*, 67 F.3d at 1101).

81. *Goldberg*, 67 F.3d at 1102–03. Because the trial court did not sufficiently warn the defendant of the dangers of self-representation, the trial court's warning was inadequate, and thus, the defendant could not have implicitly waived his right to counsel. *Id.*

82. *Id.* at 1102 ("[E]ven though there may be conduct dilatory enough to constitute a waiver by conduct but insufficient to support a pure forfeiture, we need not determine whether Goldberg's conduct in this case amounted to a waiver by conduct.") (internal citations omitted).

83. *Id.*

sufficient to support forfeiture of counsel, and the court ruled that the denial of counsel was improper.[84]

Three years later, the Third Circuit indicated that defendants can more easily forfeit the right to counsel during sentencing than at trial in *United States v. Leggett*.[85] Despite the efforts of his appointed counsel, the indigent defendant in *Leggett* was convicted of assault.[86] Upon seeing his attorney in court at his sentencing hearing, the defendant "lunged at his attorney and punched him in the head, knocking him to the ground."[87] The defendant then "straddled him and began to choke, scratch and spit on him."[88] After allowing his attorney to withdraw from the case, the trial court determined that the defendant had forfeited his right to appointed counsel.[89] The Third Circuit agreed with the trial court that the defendant's violent assault constituted "the sort of 'extremely serious misconduct'" that justifies the forfeiture of the right to counsel during sentencing.[90] However, the court specifically called attention to the fact that "the forfeiture of counsel at sentencing does not deal as serious a blow to a defendant as would the forfeiture of counsel at the trial itself."[91] The court made it clear that the fact that the defendant forfeited his right to counsel at the sentencing phase rather than at trial was important to its decision to affirm the forfeiture, but the court declined to express an "opinion as to whether [the defendant's] misconduct would have been sufficient to justify the forfeiture of counsel during the trial."[92]

Despite these rulings, the Second Circuit has taken a different view, suggesting that defendants do not forfeit their right to counsel because of a single violent incident and emphasizing that rather than utilizing forfeiture, courts should allow defendants to retain the right to counsel while employing measures to protect attorneys whenever possible.[93] In *Gilchrist v. O'Keefe*, the defendant ruptured the eardrum of his fourth appointed attorney by brutally punching him in the ear during a pre-sentencing meeting,[94] and the Second Circuit refused to overturn the state trial court's

---

84. *Id.*

85. United States v. Leggett, 162 F.3d 237, 251 n.14 (3d Cir. 1998).

86. *Id.* at 240. The defendant forced his first appointed attorney to withdraw by threatening her with physical harm. *Id.*

87. *Id.* Emergency medical personnel took the attorney to the hospital after the assault. *Id.*

88. *Id.*

89. *Id.*

90. *Id.* at 250 (quoting United States v. Goldberg, 67 F.3d 1092, 1102 (3d Cir. 1995)). The trial court did not have to conduct an evidentiary hearing on the defendant's attack because the attack took place "in full view of the district court." *Id.*

91. *Id.* at 251 n.14.

92. *Id.*

93. Gilchrist v. O'Keefe, 260 F.3d 89, 100 (2d Cir. 2001).

94. *Id.* at 90, 91 n.1.

ruling that the defendant had forfeited his right to counsel.[95] However, the court made it clear that it did not intend "to suggest that any physical assault by a defendant on counsel will automatically justify constitutionally a finding of forfeiture of the right to counsel."[96] Furthermore, the court echoed the *Leggett* court's analysis that a finding of forfeiture of counsel at trial requires higher justification than at sentencing.[97] The court emphasized that defendants who engage in a single physical attack should not have their fundamental right to counsel revoked when there is no warning that a loss of counsel could result and when it is possible that measures short of forfeiture are available to protect the safety of counsel.[98] The court stressed that, when a trial court is faced with an unruly defendant, it should first utilize safety measures such as restraints to allow attorneys to continue representation and should resort to denying the fundamental right to counsel only when the court cannot guarantee the protection of counsel.[99]

In *United States v. Thomas*, a defendant's repeated refusal to cooperate with appointed attorneys combined with the defendant's threat of physical

---

95. *Id.* The court reviewed the state court case under the deferential standard applied in habeas review because the case was before the court as a result of the defendant's appeal of the district court's denial of his petition for a writ of habeas corpus. *Id.* at 89–90; *see* 28 U.S.C. § 2254(d)(1) (2006).

96. *Gilchrist*, 260 F.3d at 100.

97. *Id.* at 99; *see* United States v. Leggett, 162 F.3d 237, 251 n.14 (3d Cir. 1998).

98. *Gilchrist*, 260 F.3d at 89. The court explained its emphasis on preserving the right to counsel:

> [H]ad this been a direct appeal from a federal conviction we might well have agreed with petitioner that the constitutional interests protected by the right to counsel prohibit a finding that a defendant forfeits that right based on a single incident, where there were no warnings that a loss of counsel could result from such misbehavior, where there was no evidence that such action was taken to manipulate the court or delay proceedings, and where it was possible that other measures short of outright denial of counsel could have been taken to protect the safety of counsel.

Id.

99. *Id.* at 100. The court emphasized that trial courts should first utilize measures short of forfeiture:

> In response to incidents of this nature, trial courts have the discretion to take intermediate steps short of complete denial of counsel, and we think that courts should exercise that discretion wherever possible and consider whether the protection of counsel can be thoroughly assured by other means—for example, keeping a defendant in restraints when meeting with counsel and during courtroom proceedings.

Id.

violence against his last attorney was grounds for forfeiture of counsel.[100] The defendant refused to communicate or cooperate with three different appointed attorneys, demanded his attorney assert unethical arguments, and threatened his last attorney with a "physical confrontation."[101] Despite a warning from the district court that continued refusal to cooperate would result in the loss of the right to counsel, the defendant continued to misbehave, and the district court ruled that the defendant had both implicitly waived and forfeited the right to counsel.[102] Upon review, the Third Circuit upheld the trial court's refusal to appoint a fifth attorney, finding that the defendant had forfeited his right to counsel as a result of his threat and his intentional efforts to delay and impede the judicial proceedings by refusing to cooperate with his attorneys.[103]

C. Forfeiture Analysis in State Courts

Many state court decisions have also recognized the concept of forfeiture and provided examples of conduct that warrants the loss of the right to counsel.[104] For example, in *Commonwealth v. Babb*, the Massachusetts Supreme Judicial Court held that a defendant who attempted to delay and disrupt the progression of his trial had forfeited his right to counsel.[105] The trial court refused to appoint new counsel to the defendant who made unsubstantiated complaints about his appointed attorney and physically assaulted his appointed attorney by thrusting a table into him and striking him on the cheek.[106] The court concluded that the defendant was "attempting to forestall his trial and disrupt the prosecution by forcing a last minute change of his court-appointed counsel."[107] Because the court believed the defendant had refused "without good cause" to proceed with his appointed attorney and would likely try to force a withdrawal of a replacement attorney, the court ruled that the defendant had forfeited the assistance of appointed counsel.[108]

In the absence of physical assault, verbal abuse and physical threats may not be enough to forfeit the right to counsel.[109] For instance, in *State v.*

---

100. United States v. Thomas, 357 F.3d 357, 363 (3d Cir. 2004).

101. *Id.* at 359–61.

102. *Id.* at 361, 363.

103. *Id.* at 363. The court noted that "[the defendant] had been verbally abusive to [his appointed attorney], tore up his correspondence, refused to cooperate in producing a witness list, and hung up on him during a telephone conversation. Further, [the defendant] attempted to force [his appointed attorney], to file several meritless, frivolous claims . . . ." *Id.*

104. *See* Commonwealth v. Means, 907 N.E.2d 646, 659–60 (Mass. 2009).

105. Commonwealth v. Babb, 625 N.E.2d 544, 546–47 (Mass. 1994).

106. *Id.* at 546. The punch bruised his face and knocked off his glasses. *Id.*

107. *Id.*

108. *Id.*

109. State v. Boykin, 478 S.E.2d 689, 692 (S.C. Ct. App. 1996).

Boykin, a South Carolina defendant verbally abused his appointed attorney and lunged after him in a waiting room before two guards could subdue him.[110] The trial court ruled that the defendant had forfeited his right to counsel because of his conduct.[111] However, the South Carolina Court of Appeals ruled that the defendant's conduct, while severe, was not extreme enough to "permanently deprive him of appointed counsel."[112] After comparing the defendant's conduct to the misconduct of the defendants in *McLeod*[113] and *Jennings*,[114] the court ruled that the defendant's conduct was not sufficiently severe to warrant the harsh sanction of forfeiture.[115]

However, a less severe assault may be grounds for forfeiture when the defendant has also repeatedly attempted to delay trial.[116] In *State v. Montgomery*, the North Carolina Court of Appeals ruled that a defendant who assaulted his attorney and repeatedly attempted to delay and disrupt his trial forfeited the right to counsel.[117] The defendant changed attorneys three times and disrupted court proceedings twice before throwing water in the face of his fourth attorney in the courtroom.[118] Noting that a defendant "may lose his constitutional right to be represented by counsel of his choice when he perverts that right to a weapon for the purpose of obstructing and delaying his trial,"[119] the court ruled that the defendant had forfeited his right to counsel, regardless of whether the trial court had warned the defendant that he could lose the right to counsel, because his disruptions and assault on his attorney were intentional attempts to delay and manipulate his trial.[120]

The California Court of Appeal has also emphasized that trial courts should exhaust all intermediate safety measures, such as verbal warnings and physical restraints, to protect attorneys from violent defendants before revoking the right to counsel.[121] When considering the case of a defendant who "head-butted" his first appointed attorney and threatened to kill his second, third, and fourth appointed attorneys,[122] the California Court of Appeal stated in *King v. Superior Court* that forfeiture of counsel "should be a court's last resort and . . . should occur only after lesser measures to

110. *Id.* at 690.

111. *Id.* The trial court subsequently appointed elbow counsel to advise the defendant. *Id.*

112. *Id.* at 692.

113. *See McLeod*, 53 F.3d at 325.

114. *See* United States v. Jennings, 855 F. Supp. 1427, 1432–33 (M.D. Pa. 1994).

115. *Boykin*, 478 S.E.2d at 691–92.

116. State v. Montgomery, 530 S.E.2d 66, 69 (N.C. Ct. App. 2000).

117. *Id.*

118. *Id.* at 67–68.

119. *Id.* at 69 (quoting State v. McFadden, 234 S.E.2d 742, 747 (N.C. 1977)).

120. *Id.* The court refused to tolerate the defendant's "purposeful conduct and tactics to delay and frustrate the orderly processes of our trial courts." *Id.*

121. King v. Superior Court, 132 Cal. Rptr. 2d 585, 588–89 (Cal. Ct. App. 2003).

122. *Id.* at 589–90.

control [the] defendant, including but not limited to a warning and physical restraints or protections, have failed."[123] In its opinion, the court insisted that forfeiture of counsel is not a necessary response to most instances of misconduct and is justified only in the most extreme situations in which intermediate measures to restrict misconduct and protect attorneys such as physical restraints have failed as a result of the defendant's intent to manipulate proceedings.[124] The court also asserted that stern warnings, rather than complete forfeiture, should occur "in instances where the misconduct does not rise to the most serious level."[125] Furthermore, the court asserted that trial courts have a duty to issue warnings to encourage defendants to conform their behavior to acceptable standards and to inform them of the consequences of additional misconduct and the risks of self-representation before denying defendants the right to counsel.[126] The court also asserted that, before depriving the defendant of the fundamental right to counsel, courts must hold an evidentiary hearing on the facts supporting forfeiture at which the defendant must be present and allowed to testify.[127] Because the defendant's interests in *King* were not adequately represented at the trial court's forfeiture hearing,[128] the court concluded that the hearing was invalid and reversed the defendants' forfeiture.[129]

Defendants who, in an effort to disrupt proceedings, repeatedly verbally abuse their attorneys (without threatening or assaulting them) can nevertheless forfeit the right to counsel.[130] In *Bultron v. State*, the Delaware Supreme Court ruled that a defendant who was intentionally abusive and offensive to his appointed attorney because he wished to receive a substitute appointed attorney forfeited the right to counsel.[131] The defendant repeatedly insulted and chastised his attorney in an effort to "force his attorney to withdraw to prevent his trial from going forward after the Superior Court had denied [his] request for substitute counsel."[132] Although

---

123. *Id.* at 588–89.

124. *Id.* at 596. Forfeiture is only appropriate "in those rare cases of extremely serious misconduct towards counsel where it is apparent that any lesser measures will be patently inadequate." *Id.*

125. *Id.*

126. *Id.*

127. *Id.* at 598. The court explained that the forfeiture hearing requires procedural due process protections: "At the hearing defendant is entitled to be present, to have the assistance of counsel, to present evidence, and to cross-examine witnesses. The court must find the facts supporting forfeiture by clear and convincing evidence, and set forth its factual findings in the record." *Id.* at 600.

128. *Id.* ("While [the defendant] nominally had counsel at the hearing, counsel not only failed to represent [the defendant's] interests, but actively argued against him and in favor of forfeiture.").

129. *Id.* at 601.

130. Bultron v. State, 897 A.2d 758, 766 (Del. 2006).

131. *Id.*

132. *Id.*

the defendant's disparaging actions "fell short of violence or threats," the court asserted that a defendant does not have to act violently to demonstrate the extremely serious misconduct necessary for forfeiture.[133] Because the defendant repeatedly harassed his appointed attorney for the purpose of obstructing and manipulating the proceedings, the court determined the defendant's conduct was sufficiently egregious to affirm the trial court's forfeiture ruling.[134]

Emphasizing the necessity of forfeiture hearings, the Massachusetts Supreme Judicial Court again addressed forfeiture in *Commonwealth v. Means* and insisted that evidentiary hearings must be held before defendants are deprived of the right to counsel due to misconduct.[135] The case concerned a defendant who sent a blood-stained letter to his attorney threatening to harm him and his family.[136] Noting that forfeiture of counsel requires "extraordinary circumstances,"[137] the court asserted that the appropriateness of forfeiture may be determined by considering a number of factors: whether the defendant has had more than one appointed attorney, whether the defendant will be deprived of counsel at trial or at secondary phase of the proceedings rather than at trial, whether the defendant has assaulted or threatened his attorney with violence, and whether intermediate measures less onerous than forfeiture are insufficient to protect counsel.[138] The court also stated that before a judge determines whether the defendant's conduct was extreme enough to warrant the severe sanction of forfeiture of counsel, it is essential that trial courts hold a hearing in which defendants, represented by counsel, have the opportunity to offer evidence and question witnesses to demonstrate that forfeiture is not warranted.[139] Because the trial court did not hold a proper forfeiture hearing,[140] the court ruled that the defendant did not forfeit his right to counsel and remanded the case for a new trial.[141]

---

133. *Id.* (citing United States v. McLeod, 53 F.3d 322, 325–26 (11th Cir. 1995)).

134. *Id. See generally* WAYNE R. LAFAVE ET AL., CRIMINAL PROCEDURE § 11.3(c), at 582 (3d ed. 2000) ("[T]he state's interest in maintaining an orderly trial schedule and the defendant's negligence, indifference, or possibly purposeful delaying tactic, combined to justify a forfeiture of defendant's right to counsel in much the same way that the defendant's assault upon his counsel can result in his loss of representation . . . .").

135. Commonwealth v. Means, 907 N.E.2d 646, 652 (Mass. 2009).

136. *Id.* at 653.

137. *Id.* at 659.

138. *Id.* at 659–61.

139. *Id.* at 652, 664.

140. *Id.* at 662–63. The court deemed the lower court's hearing inadequate "[b]ecause no notice had been given that the judge was considering forfeiture, there was no evidence presented as to the defendant's psychological condition, the circumstances that led up to the threatening letter, or even defense counsel's willingness to continue as counsel . . . ." *Id.*

141. *Id.* at 664.

D. Forfeiture Analysis in Tennessee

Addressing the issue of forfeiture for the first time in *State v. Carruthers*, the Tennessee Supreme Court acknowledged the differences between waiver, implicit waiver, and forfeiture of counsel[142] and ruled that a defendant who had forced the withdrawal of six appointed attorneys had both implicitly waived and forfeited his right to counsel.[143] The court recognized that forfeiture can occur when a defendant "engages in extremely serious misconduct" despite the fact that "defendant was not warned of the potential consequences of his or her actions or the risks associated with self-representation" and "regardless of the defendant's intent to relinquish the right."[144] Refusing to appoint another attorney, the trial court forced the defendant to represent himself at trial and during sentencing after six previously appointed attorneys withdrew because of threats of physical harm and verbal abuse by the defendant.[145] Asserting that an "indigent criminal defendant may implicitly waive or forfeit the right to counsel by utilizing that right to manipulate, delay, or disrupt trial proceedings," the Tennessee Supreme Court ruled that the defendant had both implicitly waived and forfeited his right to counsel.[146] The court explained that the defendant implicitly waived his right to counsel by "glaring at" and sending harassing letters to his fifth attorney after receiving a warning that further misconduct would result in the loss of the right to counsel.[147] Furthermore, the court concluded that the defendant also forfeited the right to counsel by repeatedly demanding substitute appointed counsel, threatening appointed attorneys with physical violence, and making baseless allegations against his attorney in order to obstruct and manipulate the proceedings.[148] Because the defendant purposefully "sabotaged his relationship with each successive attorney with the obvious goal of delaying and disrupting the orderly trial of the case," the court held that the defendant's conduct was egregious enough to also support a finding of forfeiture independent of his implicit waiver of counsel.[149] Thus, according to *Carruthers*, a defendant can lose the right to counsel in Tennessee even in the absence of a judge's warning.[150]

The Tennessee Court of Criminal Appeals addressed forfeiture in *State v. Small* and held that a defendant who violently assaults his attorney can

142. State v. Carruthers, 35 S.W.3d 516, 546–49 (Tenn. 2000).
143. *Id.* at 549–50.
144. *Id.* at 548.
145. *Id.* at 543–45.
146. *Id.* at 549–50.
147. *Id.* at 549.
148. *Id.* at 550.
149. *Id.*
150. *Id.*

forfeit the right to counsel.[151] In *Small*, the defendant punched his attorney in the face twice, bloodying his lip.[152] The trial court ruled that this assault was sufficient to warrant the sanction of forfeiture of counsel.[153] Citing *Leggett*, the court of appeals held that the defendant's violent misconduct was extreme and egregious enough to justify the trial court's finding of forfeiture.[154]

In *State v. Willis*, the Tennessee Court of Criminal Appeals again confronted a forfeiture case, this time holding that a defendant who repeatedly tries to disrupt and manipulate trial proceedings by exploiting the right to counsel can forfeit that right.[155] The defendant in *Willis* filed fifty-five unfounded complaints against his first two attorneys, refused to answer questions or provide information for his attorneys, and repeatedly requested that the court appoint substitute counsel.[156] After seven changes of counsel and numerous warnings from the court, the trial court concluded that the defendant had "egregiously manipulated the constitutional right to counsel resulting in delay" and thus had forfeited the right to counsel.[157] Relying on the Tennessee Supreme Court's decision in *Carruthers*,[158] the appeals court affirmed the ruling of the trial court for two reasons.[159] First, the court stated that the defendant had implicitly waived his right to counsel because he "persisted in intentional conduct that prompted the disqualification of counsel" despite several warnings from the judge about the consequences of further manipulative conduct and the dangers of self-representation in a capital trial.[160] In addition, the court found that the sanction of forfeiture was appropriate because the defendant repeatedly made baseless allegations against his attorneys and refused to cooperate with his attorneys in order to obstruct and delay the proceedings.[161]

---

151. State v. Small, No. W2007-01723-CCA-R3-CD, 2009 WL 331323, at *1 (Tenn. Crim. App. Feb. 10, 2009).

152. *Id.* at *2. The defendant argued that the attorney provoked him into striking him by uttering a racial slur, but the trial court "accredit[ed] the testimony of counsel" that no such racial slur was uttered. *Id.* at *3.

153. *Id.* at *3.

154. *Small*, 2009 WL 331323, at *4.; *see* United States v. Leggett, 162 F.3d 237, 250 (3d Cir. 1998).

155. State v. Willis, 301 S.W.3d 644, 652 (Tenn. Crim. App. 2009).

156. *Id.* at 646–49.

157. *Id.* at 649.

158. State v. Carruthers, 35 S.W.3d 516, 549 (Tenn. 2000).

159. *Willis*, 301 S.W.3d at 651.

160. *Id.* at 652.

161. *Id.*

IV. THE DEFENDANT'S CONDUCT WAS NOT SUFFICIENTLY EGREGIOUS TO JUSTIFY THE EXTREME SANCTION OF FORFEITURE OF THE RIGHT TO COUNSEL

In *State v. Holmes*, the Tennessee Supreme Court held that the defendant's single assaultive act and ambiguous threat did not rise to the level of "extremely serious misconduct" necessary to warrant forfeiture of the right to counsel.[162] The court based its decision on the particular circumstances of the case: (1) forfeiture would have denied the defendant the assistance of counsel at trial rather than at a later secondary proceeding; (2) the defendant was not trying to delay or disrupt proceedings; (3) his attack resulted in no bodily injury to his attorney; (4) his attack consisted of a single assault on his first and only appointed attorney; and (5) the trial court could have utilized measures other than forfeiture to protect his attorney's safety.[163]

Justice Clark wrote the court's unanimous opinion and began her analysis by recognizing that while criminal defendants have the fundamental right to counsel at trial, the right is not absolute.[164] Citing the court's first analysis of forfeiture of counsel in *Carruthers*, the court stated that a defendant can forfeit the right to counsel upon a showing of "'extremely serious misconduct'" or an "'egregious manipulation' of the right to counsel" to intentionally "'delay, disrupt, or prevent the orderly administration of justice.'"[165] In addition, the court asserted that trial courts must hold "an evidentiary hearing at which the defendant is present" and allowed to testify to determine if a defendant has behaved in a manner that warrants a finding of forfeiture.[166] The court also enumerated specific factors that trial courts should consider when deciding if forfeiture is appropriate: "(1) whether the defendant has had more than one appointed counsel; (2) the stage of the proceedings . . . ; (3) violence or threats of violence against appointed counsel; and (4) measures short of forfeiture have been or will be unavailing."[167]

The court next reviewed its analysis of forfeiture in *Carruthers* in which the court distinguished waiver, implicit waiver, and forfeiture of the right to counsel.[168] The court reiterated that implicit waiver occurs when the defendant's misconduct persists "*after* he has been made aware that his

---

162. State v. Holmes, 302 S.W.3d 831, 847–48 (Tenn. 2010) (quoting *Carruthers*, 35 S.W.3d at 548).

163. *Id.* at 848.

164. *Id.* at 833, 838.

165. *Id.* at 838 (quoting *Carruthers*, 35 S.W.3d at 548, 550).

166. *Id.* at 838–39 (citing *Means*, 907 N.E.2d at 662 (Mass. 2009)). Hearings are not necessary if the defendant's misconduct occurred in plain view of the court. *Id.* at 839 n.6 (citing United States v. Leggett, 162 F.3d 237, 250 (3d Cir. 1998)).

167. *Id.* at 839 (citing *Means*, 907 N.E.2d at 659–61).

168. *Id.* at 840 (citing *Carruthers*, 35 S.W.3d at 546, 548–49).

continued misbehavior will result in the dangers and disadvantages of proceeding pro se."[169] Furthermore, the court asserted that forfeiture is appropriate when a defendant "engages in extremely serious misconduct . . . even though the defendant was not warned of the potential consequences of his or her actions or the risks associated with self-representation."[170]

The court then distinguished *Carruthers* from the instant case, noting that the *Carruthers* defendant had both implicitly waived the right to counsel by continuing to be disruptive even "after being warned by the trial court that he would lose his attorney if his misconduct continued" and forfeited the right to counsel by manipulating the right "with the obvious goal of delaying and disrupting the orderly trial of the case."[171] The court explained that while "the facts and circumstances" of *Carruthers* supported a finding of either implicit waiver or forfeiture, a finding of implicit waiver was not proper in the instant case.[172] The trial court in the immediate case never warned the defendant of the consequences misconduct could entail.[173] Accordingly, the court concluded, the trial court could have denied the right to counsel only on the basis of forfeiture, not by implied waiver.[174] Furthermore, unlike the defendant in *Carruthers* who purposely obstructed proceedings by forcing the withdrawal of several attorneys through unfounded allegations and threats, the instant defendant engaged in "no pattern of abusive behavior or conduct aimed at delaying the proceedings," but only a single assault which caused no injury and an ambiguous verbal threat.[175] Thus, the court reasoned that because the conduct of the defendants was so different, *Carruthers* does not provide guidance on whether the instant defendant's conduct was egregious enough to justify forfeiture.[176]

With no other Tennessee Supreme Court cases addressing the issue of forfeiture,[177] the court sought guidance from decisions of other jurisdictions involving a defendant's forfeiture of the right to counsel due to a physical attack to determine whether the defendant's conduct in the instant case warranted the sanction of forfeiture.[178] The court first reviewed relevant federal cases from the Second and Third Circuits[179] and then considered

169. *Id.*

170. *Id.* (quoting *Carruthers*, 35 S.W.3d at 548).

171. *Id.* (quoting *Carruthers*, 35 S.W.3d at 549–50).

172. *Id.* at 840–41 (citing *Carruthers*, 35 S.W.3d at 549).

173. *Id.* at 841.

174. *Id.*

175. *Id.* at 841 (referencing *Carruthers*, 35 S.W.3d at 550).

176. *Id.*

177. *Id.* at 839.

178. *Id.* at 841.

179. *Id.* at 841–43 (citing Gilchrist v. O'Keefe, 260 F.3d 87, 94 (2d Cir. 2001); United States v. Leggett, 162 F.3d 237 (3d Cir. 1998); United States v. Jennings, 855 F. Supp. 1427 (M.D. Pa. 1994)).

state court cases from the California Court of Appeal, the North Carolina Court of Appeals, the South Carolina Court of Appeals, and the Massachusetts Supreme Judicial Court.[180]

After this lengthy analysis of similar cases, the court concluded that the fundamental right to counsel is of such importance that forfeiture of the right without warning is appropriate only in response to the most extreme behavior.[181] To illustrate its position, the court quoted the Massachusetts Supreme Judicial Court's analysis of forfeiture in *Commonwealth v. Means* in which it stated that forfeiture is such a severe sanction that it should be employed only in "extraordinary circumstances" as a "last resort in response to the most grave and deliberate misconduct."[182] The court also endorsed the *Gilchrist* court's opinion that a defendant should not forfeit "his right to counsel at trial on the basis of a single incident of physical violence unless the violence was extreme" and (1) the trial court had previously warned the defendant "that he could lose the right to counsel for such behavior," (2) evidence established that the "defendant engaged in the violence in order to manipulate the court or delay the proceedings," or (3) intermediate measures could not be utilized to "protect the safety of counsel."[183] The court also briefly addressed the Tennessee Court of Criminal Appeals' decision in *State v. Small*, a case similar to the instant case involving a defendant who punched his attorney in the face,[184] but the court expressed no opinion on the trial court's decision to deny the right to counsel at sentencing or the Court of Criminal Appeals' refusal to reverse.[185]

Turning to the instant case, the court compared the actions of the defendant to the standards for forfeiture discussed in the previous cases.[186] The court implied that the ambiguity of the defendant's threat reduced its severity as evidenced by the fact that Counsel did not initially feel the need to report the threat to the trial judge or request that the defendant be handcuffed at the next meeting.[187] Moreover, the court stressed that there was no evidence to suggest that the defendant took any action purposely designed "to delay or disrupt the proceedings."[188] In fact, the court observed

---

180. *Id.* at 843–46 (citing King v. Superior Court, 132 Cal. Rptr. 2d 585 (Cal. Ct. App. 2003); Commonwealth v. Babb, 625 N.E.2d 544 (Mass. 1994); State v. Montgomery, 530 S.E.2d 66 (N.C. Ct. App. 2000); State v. Boykin, 478 S.E.2d 689 (S.C. Ct. App. 1996)).

181. *Id.* at 846.

182. *Id.* at 847 (quoting Commonwealth v. Means, 907 N.E.2d 646, 658–60 (Mass. 2009)).

183. *Id.* (citing *Gilchrist*, 260 F.3d at 89).

184. *Id.* at 846 n.9 (citing State v. Small, No. W2007-01723-CCA-R3-CD, 2009 WL 331323, at *2 (Tenn. Crim. App. Feb. 10, 2009)).

185. *Id.* at 847 n.9.

186. *Id.* at 847.

187. *Id.*

188. *Id.*

that the defendant never requested substitute counsel, and Counsel was the first and only attorney appointed.[189] In addition, the court noted that the trial court did not attempt to use intermediate protective measures before resorting to forfeiture and emphasized that there was no evidence to suggest that "less onerous corrective measures, such as shackling [the defendant] during meetings with his lawyer, would not have been adequate to insure his lawyer's future safety."[190]

Addressing the defendant's assault, the court asserted that although the defendant's pushing of the attorney's glasses with his finger constituted serious misconduct, the court refused to characterize it as the sort of "extremely serious misconduct" necessary for a finding of forfeiture.[191] The court insisted that whether a defendant's misconduct is egregious enough to justify the extreme sanction of total forfeiture of counsel at trial without warning "depends upon the particular facts and circumstances of the attack at issue."[192] Emphasizing the harshness of forfeiture, the court ruled that in light of the particular circumstances of the case, the defendant's conduct was not extreme enough to support a finding of forfeiture.[193] Accordingly, the court ruled that the trial court erred in finding that the defendant forfeited his right to counsel, and the court reversed the defendant's conviction and remanded the case to the trial court.[194]

The court concluded by outlining steps trial courts should take when faced with a defendant who has allegedly attacked his appointed attorney.[195] The court established that the trial court should first hold an evidentiary hearing in which the defendant is present and permitted to testify.[196] The trial court should make findings of fact based on the evidence presented at the hearing, and from those facts, the trial court should determine whether the defendant's conduct was extreme enough to support a finding of forfeiture or implicit waiver of counsel.[197] When determining if forfeiture is appropriate, the trial court should consider: (1) "the stage of the proceedings;" (2) how many previous lawyers have withdrawn as a result of the defendant's misconduct; (3) whether the defendant has been warned that forfeiture of counsel can result from misconduct; (4) whether the

189. *Id.*

190. *Id.*

191. *Id.* at 847–48 (quoting State v. Carruthers, 35 S.W.3d 516, 548 (Tenn. 2000)).

192. *Id.* at 847. The court explained that physical assault is not always required to support a finding of forfeiture, and threats could be sufficient for forfeiture when "a defendant repeatedly threatens harm to his lawyer and/or his lawyer's family and it is apparent that the defendant has the ability to deliver on his threats." *Id.* at n.10.

193. *Id.* at 847–48 (listing five specific circumstances of the case justifying the court's ruling).

194. *Id.* at 848. The court said the appointed attorney "should be allowed to withdraw if requested." *Id.*

195. *Id.*

196. *Id.*

197. *Id.*

defendant's actions were part of an intentional pattern of behavior meant to obstruct or manipulate the proceedings; (5) "the degree of violence" and the severity of any resulting injury; and (6) whether measures other than forfeiture could be used to adequately protect counsel.[198] Upon determining that the defendant's conduct was not extreme enough to support forfeiture, the trial court should: (1) appoint a new attorney if the prior attorney withdrew; (2) warn the defendant that the loss of counsel could result from future misbehavior; (3) inform the defendant about the risks of self-representation; and (4) utilize measures to protect the new attorney from the defendant's future misconduct.[199]

V. IMPLICATIONS OF *STATE V. HOLMES*

The Tennessee Supreme Court's decision in *State v. Holmes* demonstrates the importance the court places on the Sixth Amendment right to counsel as well as the court's belief that the right to counsel should not be easily waived or forfeited.[200] The ruling does not depart from the court's analysis of forfeiture in *Carruthers* and reflects the trend in other jurisdictions of preserving the defendant's right to counsel when the defendant's misconduct does not include extreme violence, overt threats, or intentional manipulation of the right to counsel to delay or disrupt proceedings.[201] The court's suggested procedure for handling defendants who assault their attorneys reinforces the court's directive that trial courts should take all the facts and circumstances into account when determining if misconduct warrants forfeiture.[202]

Although few Tennessee cases provide guidance on the issue of forfeiture, many cases from other jurisdictions have confronted forfeiture in similar contexts.[203] And although these cases involve facts and circumstances that differ from *Holmes*, the courts hearing these cases generally approach forfeiture in a consistent manner: by examining the defendant's misconduct and balancing the defendant's fundamental right to counsel against the societal interests of judicial efficiency and the safety of appointed attorneys.[204] The Tennessee Supreme Court applied this approach in *Holmes* and correctly concluded that, on balance, the Defendant's right

198. *Id.*

199. *Id.*

200. *Id.* at 846; *see also* C. Allen Parker, Jr., Note, *Proposed Requirements for Waiver of the Sixth Amendment Right to Counsel*, 82 COLUM. L. REV. 363, 365 (1982) ("[T]he broader protections of the sixth amendment right are most effectively ensured through stringent waiver requirements . . . .").

201. *Holmes*, 302 S.W.3d at 846–47; *see* Commonwealth v. Means, 907 N.E.2d 646, 659–60 (Mass. 2009); State v. Carruthers, 35 S.W.3d 516, 546–49 (Tenn. 2000).

202. *Holmes*, 302 S.W.3d at 848.

203. *Id.* at 839–41.

204. *See* Parker, *supra* note 54, at 1207.

to counsel at trial was more important than eliminating the risk to attorney safety by forcing the Defendant to represent himself because intermediate measures such as handcuffs would likely be sufficient to prevent further harm.[205]

In its analysis, the court correctly concluded that, while the defendant can both implicitly waive and forfeit the right to counsel,[206] the instant Defendant's misconduct should be analyzed as a forfeiture and not an implicit waiver because the trial court did not warn the Defendant that further misconduct would result in forfeiture of counsel.[207] Applying its own standards for forfeiture from *Carruthers* and the standards from the decisions of other jurisdictions, the court correctly determined that the Defendant's ambiguous threat and finger poke did not constitute the extreme misconduct required to justify the severe sanction of forfeiture of the right to counsel.[208] The Defendant's misconduct was not as violent or injurious as the behavior demonstrated in cases such as *Leggett* and *Jennings*, and unlike the defendant in *Carruthers*, there was no pattern of behavior to suggest that the defendant was trying to manipulate his right to counsel in order to obstruct or delay proceedings.[209] Furthermore, no other attorneys had represented the Defendant, and there was no indication that measures such as handcuffing the defendant would not have been sufficient to protect counsel while still allowing the defendant to avoid forfeiture at the critical trial stage.[210]

The court properly ruled that the Defendant's conduct was not egregious enough to warrant forfeiture of the fundamental right to counsel, but it failed to give a clear definition of "extremely serious conduct" and instead pronounced that whether an assault meets this standard is contingent upon the facts and circumstances of the particular case.[211] The lack of a decisive standard for extremely serious conduct makes it difficult for trial court judges to determine whether conduct is sufficiently egregious to justify forfeiture. The court's recommended forfeiture procedure provides some assistance in this regard, but it includes so many factors to consider that it is unclear which factors should be critical or determinative in the eyes of the trial court.

The scope of the *Holmes* ruling is fairly limited because the court's holding was contingent upon the facts and circumstances of the particular

---

205. *Holmes*, 302 S.W.3d at 847–48.

206. *See* Parker, *supra* note 54, at 1212 ("Courts should . . . recognize that forfeiture of the right to counsel may occur as a matter of law.").

207. *Holmes*, 302 S.W.3d at 841.

208. *Id.* at 847–48; *see* State v. Carruthers, 35 S.W.3d 516, 546–49 (Tenn. 2000).

209. *Holmes*, 302 S.W.3d at 847–48; *see also* United States v. Leggett, 162 F.3d 237, 240 (3d Cir. 1998); United States v. Jennings, 855 F. Supp. 1427, 1432–33 (M.D. Pa. 1994); *Carruthers*, 35 S.W.3d at 550.

210. *Holmes*, 302 S.W.3d at 847–48.

211. *Id.*

case.[212] Nevertheless, the decision will affect forfeiture analysis in Tennessee, especially when the defendant has engaged in similar assaultive misconduct. By emphasizing the importance of counsel at trial,[213] the *Holmes* decision indicates that defendants can engage in more severe misconduct during and before trial without forfeiting the right to counsel than they can after the verdict. Furthermore, the ruling implies that a defendant's attack must physically injure someone before it will constitute extreme misconduct.[214] Moreover, the court's holding suggests that a trial court cannot revoke the right to counsel of an indigent defendant who vaguely threatens or simply humiliates appointed counsel unless the defendant clearly threatens to harm the attorney and is reasonably capable of carrying out the threat.[215] Because intermediate measures short of forfeiture will usually be sufficient to protect the safety of appointed attorneys, the ruling also suggests that a finding of forfeiture will be difficult to justify if the trial court has not first utilized measures short of forfeiture to protect defendants.[216]

By offering suggested procedures for courts to follow when faced with a defendant who has allegedly assaulted his attorney,[217] *Holmes* has implications beyond its holding. For instance, trial courts now have a list of factors to consider when determining whether a defendant's misconduct is sufficient to warrant forfeiture of counsel.[218] Although trial courts now have less latitude to use their own judgment when deciding whether a defendant has forfeited the right to counsel, the list of factors promotes judicial efficiency and the consistent application of the law by providing a standard rubric for evaluating a defendant's misconduct. Furthermore, attorneys who have allegedly been assaulted by their clients can now more easily withdraw from the case.[219] In addition, the decision will encourage trial judges to both formally warn defendants and employ measures to protect appointed attorneys, such as handcuffs or other physical restraints, at the first sign of misconduct from a defendant. Increasing these preventative measures will reduce defendant misconduct and increase attorney safety.

VI. CONCLUSION

The decision in *State v. Holmes* represents the Tennessee Supreme Court's commitment to preserving the crucially important right to counsel

212. *Id.* at 848.
213. *Id.*
214. *Id.*
215. *Id.* at 847 n.10.
216. *Id.* at 848.
217. *Id.*
218. *Id.*
219. *Id.* The court explicitly stated that the appointed attorney "should be allowed to withdraw if requested." *Id.*

whenever a defendant's misconduct is not extremely serious and deliberate.[220] Following the precedent set by the court in *Carruthers* and by courts in other jurisdictions, the court correctly concluded that in light of the facts and circumstances of the case, the defendant's ambiguous threat and finger punch did not constitute the extremely serious misconduct necessary to warrant forfeiture of the fundamental right to counsel.[221] Because means of protecting counsel other than forfeiture were available and the defendant's single assault against his only attorney did not cause injury, forfeiture of counsel at trial without warning was not appropriate.[222] While the right to counsel is fundamental, courts must be careful not to overzealously uphold this right at the expense of judicial efficacy and attorney safety. Although the court's suggested forfeiture analysis provides valuable guidance to trial courts confronting defendants who have engaged in assaultive misconduct, the question as to what exactly constitutes extremely serious misconduct remains unanswered.

SCOTT M. MCLEOD

220. *Id.* at 846.
221. *Id.* at 847–48.
222. *Id.* at 848.

CIVIL PROCEDURE—INTERVENTION—NONSOVEREIGN ENTITIES IN EQUITABLE APPORTIONMENT ACTIONS INVOKING ORIGINAL JURISDICTION

South Carolina v. North Carolina, 130 S. Ct. 854 (2010).

I. INTRODUCTION

The riparian[1] eastern states, once known for their "'bountiful rain and plentiful lakes and rivers,'" were seemingly impervious to drought and water scarcity that was common throughout the western states.[2] After World War II, the demand for water increased throughout the country as a result of rapid economic and population growth, highlighting the shortcomings of eastern riparianism.[3] In South Carolina, dramatic population increases, severe droughts, and increased interstate competition for water have made the resource an "object of combative courting" in past decades.[4] Historically, South Carolina and North Carolina shared the waters of the Catawba River[5] without incident.[6]

---

1. The doctrine of riparianism assumes that water is bountiful and readily available to all. Alyssa S. Lathrop, Comment, *A Tale of Three States: Equitable Apportionment of the Apalachicola-Chattahoochee-Flint River Basin*, 36 FLA. ST. U. L. REV. 865, 881 (2009). According to this doctrine, "'all uses [of water], regardless of when they began, are allowed provided they do not unreasonably interfere with other uses.'" *Id.* (quoting C. Grady Moore, *Water Wars: Interstate Water Allocation in the Southeast*, 14 NAT'L RESOURCES & ENV'T 5, 6 (1999)).

The arid western states follow the doctrine of prior appropriation. *Id.* at 880–81. Under this doctrine, "once a water user has acquired a water right, his or her right is superior to any water uses that arise later." *Id.* at 881 (quoting Moore, *supra*, at 6).

2. *Id.* at 866 (quoting J.B. Ruhl, *Equitable Apportionment of Ecosystem Services: New Water Law for a New Water Age*, 19 J. LAND USE & ENVTL. L. 47, 47 (2003)).

3. Catherine D. Little, *Eastern Water Law: Less Water, More Change*, 39 TRENDS (A.B.A. Sec. of Environmentt, Energy, & Resources Newsletter), March/April 2008, at 8.

4. J. Blanding Holman IV, *The Advent of Modified Riparianism in South Carolina*, 16 SE. ENVTL. L.J. 291, 293 (2008).

5. The river "originates in the North Carolina mountains and winds 225 miles into South Carolina, crossing the border at Lake Wylie." Brief of the State of South Carolina in Support of Its Motion for Leave to File Complaint at 1, South Carolina v. North Carolina, 130 S. Ct. 854 (2010) (No. 138). American Rivers, a Washington, D.C. conservation group, rated the Catawba River as the most endangered river in the United States in 2008. Bruce Henderson & Adam O'Daniel, *Fast-Growing Water Demand, Drought, Faulty Policies Lead to Ominous Ranking*, HERALD, Apr. 17, 2008, at A1, *available at* http://www.heraldonline.com/2008/04/17/494198/group-designates-catawba-asnations.html.

6. Brief of the State of South Carolina in Support of Its Motion for Leave to File Complaint, *supra* note 5, at 1. Both North Carolina and South Carolina depend on the river

In 1991, North Carolina enacted a statute[7] allowing transfers of up to 2 million gallons of water per day ("mgd") from the Catawba River basin without prior authorization from state officials.[8] The statute also allows for transfers of more than 2 mgd from the river basin by obtaining a permit from the North Carolina Environmental Management Commission.[9] South Carolina claims that these transfers exceed North Carolina's equitable share of the river, thus depriving South Carolina of its equitable share.[10] In June 2007, South Carolina filed a motion for leave to file a complaint with the United States Supreme Court, invoking the Court's original jurisdiction under 28 U.S.C. § 1251.[11] The Court granted leave for South Carolina to file its complaint later that year.[12] The complaint sought a decree equitably apportioning the Catawba River between South Carolina and North Carolina, enjoining North Carolina from authorizing transfers of water from the river that exceed the state's equitable share, and declaring the North Carolina transfer statute invalid if the transfers exceed its equitable share.[13]

After granting leave to South Carolina to file its complaint, two nonstate entities, the Catawba River Water Supply Project ("CRWSP")[14]

---

for "hydroelectric power, economic development and commerce, and recreation in an area encompassing more than 1.5 million people and the Charlotte metropolitan area, which spans both States." *Id.*

7. *See* N.C. GEN. STAT. § 143-215.22G(1)(h) (2009) (defining the Catawba River Basin as an applicable river basin for the purposes of the statute); N.C. GEN. STAT. § 143-215.22L(a)(1) (2009) (authorizing the transfers).

8. South Carolina v. North Carolina, 130 S. Ct. 854, 859 (2010).

9. *Id.* South Carolina claimed that "North Carolina has issued at least two such permits, one to Charlotte for the transfer of up to 33 mgd, and one to the North Carolina cities of Concord and Kannapolis for the transfer of 10 mgd." *Id.*

10. *Id.*

11. *Id.*; *see also* U.S. CONST. art. III, § 2; 28 U.S.C. § 1251(a) (2006) (stating that "[t]he Supreme Court shall have original and exclusive jurisdiction of all controversies between two or more States").

12. *South Carolina*, 130 S. Ct. at 859; South Carolina v. North Carolina, 552 U.S. 804 (2007).

13. Complaint at 10, South Carolina v. North Carolina, 130 S. Ct. 854 (2010) (No. 138) [hereinafter Complaint]. The Court noted that the Complaint was silent as to a minimum flow of water to satisfy South Carolina's requirements. *South Carolina*, 130 S. Ct. at 859. However, South Carolina offered information from the results of a "'multi-stakeholder negotiation process'" between the Federal Energy Regulatory Commission ("FERC"), Duke Energy, and other groups from both states. *Id.* (quoting Complaint at 14). The parties to this negotiation agreed that South Carolina should receive a continuous flow of no less than 711 mgd. *Id.* (quoting Complaint at 14). However, in analyzing the feasibility of this agreed upon figure, the Court observed that "the Catawba River—even in its natural state—often would not deliver into South Carolina" a minimum average flow of 711 mgd based on estimates from Duke Energy. *Id.* at 860 (quoting Complaint at 8, 16).

14. Reply Brief in Support of the Motion of the Catawba River Water Supply Project for Leave to Intervene at 3, South Carolina v. North Carolina, 130 S. Ct. 854 (2010) (No. 138). The CRWSP is a publicly-owned "joint venture by the Lancaster County Water and

and Duke Energy Carolinas, LLC ("Duke Energy"),[15] filed motions for leave to intervene as parties.[16] In the CRWSP's motion to intervene as a party-defendant, it "assert[ed] its interest as a 'riparian user of the Catawba River' and claim[ed] that this interest was not adequately represented because of the CRWSP's 'interstate nature.'"[17] Duke Energy sought leave to intervene and file an answer claiming that neither State could adequately represent its "particular amalgam of federal, state and private interests."[18] South Carolina opposed the motions of both the nonstate entities, and the Court appointed a Special Master, referring the matter and the motions to her.[19]

One month after referring the motions to the Special Master, the city of Charlotte also filed for leave to intervene as a party-defendant.[20] Charlotte asserted that North Carolina could not adequately represent the city's interests because "the State was duty bound to represent the interests of all North Carolina users of the Catawba River's water, including users whose

---

Sewer District ("LCWSD") in South Carolina and Union County ("UC") in North Carolina." *Id.*

15. Duke Energy Carolinas, LLC's Reply Brief in Support of Motion to Intervene and File Answer at 2, South Carolina v. North Carolina, 130 S. Ct. 854 (2010) (No. 138). Duke Energy impounds water from the Catawba River in eleven dams and reservoirs in both states to generate and provide hydroelectric power throughout the region, and to determine the flow of the river. *Id.* Duke Energy's ability to regulate the river is governed by a fifty-year FERC license. *South Carolina*, 130 S. Ct. at 860 n.1. In 2006, Duke Energy coordinated a negotiation process that resulted in a Comprehensive Relicensing Agreement ("CRA") signed by seventy entities from both states. *Id.* at 860. This CRA determines the terms and conditions for Duke Energy's application to renew its FERC license. *Id.* Currently, "Duke Energy is operating under a temporary extension of its 50-year FERC license, which expired in 2008, and the CRA represents Duke Energy's investment in a new 50-year license." *Id.* at 866 n.7.

16. *Id.* at 860.

17. *Id.*

18. *Id.*

19. *South Carolina*, 130 S. Ct. at 860. Acting as a trial court in original jurisdiction cases, the Court customarily delegates its authority to Special Masters. Michael Coenen, Comment, *Original Jurisdiction Deadlocks*, 118 YALE L.J. 1003, 1009 (2009) (citation omitted). Special Masters are appointed to manage the development of the case, providing recommendations for the Court. *Id.* For a critical analysis of the appointment of Special Masters, *see generally* Anne-Marie C. Carstens, *Lurking in the Shadows of Judicial Process: Special Masters in the Supreme Court's Original Jurisdiction Cases*, 86 MINN. L. REV. 625, 628 (2002).

20. *South Carolina*, 130 S. Ct. at 860. With a population of more than 800,000, Charlotte is the largest municipality and "largest provider of water supply and wastewater treatment services in the Catawba River Basin." Reply Brief for the City of Charlotte, North Carolina at 1, South Carolina v. North Carolina, 130 S. Ct. 854 (2010) (No. 138). The city holds a permit authorizing the largest transfer from the river identified in South Carolina's complaint, 33 mgd.

interests were not aligned with Charlotte's."[21] South Carolina again opposed the motion to intervene, and the Court referred it to the Special Master.[22]

With all three motions to intervene before her, the Special Master held a hearing and ruled on the motions.[23] Using the Court's "appropriate standard" for a nonstate entity's motion to intervene in an original jurisdiction action, the Special Master "distilled" a rule governing the motions.[24] After applying this formulated rule to the motions before her, the Special Master found that "each proposed intervenor had a sufficiently compelling interest to justify intervention."[25] Accordingly, the Special Master recommended that the Court grant all three motions to intervene.[26] Upon South Carolina's request, the Special Master issued a First Interim Report ("Report") presenting her findings and decisions.[27] South Carolina then presented exceptions to this Report.[28] On original jurisdiction, the United States Supreme Court, *held*, overruled as to the CRWSP and Duke Energy, and sustained as to the city of Charlotte, North Carolina. Because

---

21. *Id.* at 860–61 (citation omitted). Charlotte also mentioned its interest in protecting the terms of the above-mentioned CRA that North Carolina had not signed due to "conflicting duties under § 401 of the Clean Water Act." *Id.* at 860 n.2. The Court noted, however, that this argument was not dispositive in the recommendations of the Special Master, and Charlotte had not reasserted the argument. *Id.* As such, the Court did not consider the argument in its analysis. *Id.*

22. *Id.* at 861 (citation omitted).

23. *Id.*

24. *Id.* (citation omitted); *see* New Jersey v. New York, 345 U.S. 369, 373 (1953) (illustrating the basis for the rule the special master developed). The Special Master's rule states:

> Although the Court's original jurisdiction presumptively is reserved for disputes between sovereign states over sovereign matters, nonstate entities may become parties to such original disputes in appropriate and compelling circumstances, such as where the nonstate entity is the instrumentality authorized to carry out the wrongful conduct or injury for which the complaining state seeks relief, where the non-state entity has an independent property interest that is directly implicated by the original dispute, where the non-state entity otherwise has a "direct stake" in the outcome of the action within the meaning of the Court's cases discussed above, or where, together with one or more of the above circumstances, the presence of the nonstate entity would advance the "full exposition" of the issues.

South Carolina, 130 S. Ct. at 861 (citation omitted).

25. *South Carolina*, 130 S. Ct. at 861.

26. *Id.* Along with her recommendations on the motions, the Special Master also "rejected South Carolina's proposal to limit intervention to the remedy phase of this litigation . . . ." *Id.*

27. *Id.*

28. *Id.*

the CRWSP and Duke Energy are able to show "compelling interests" that would not be adequately represented by a sovereign party, they are permitted to intervene as nonsovereign entities in an original jurisdiction action. *South Carolina v. North Carolina*, 130 S. Ct. 854 (2010).

II. DEVELOPMENT OF NONSOVEREIGN INTERVENTION IN EQUITABLE APPORTIONMENT ACTIONS INVOKING THE COURT'S ORIGINAL JURISDICTION

Along with the increased development and attendant increased demand for water, there has been a significant change in the former riparian nature of the eastern states.[29] With the increasing competition for water over the past decades, the Court's original jurisdiction docket has grown to include struggles in both the arid western states and the drought-ridden eastern states.[30] *South Carolina v. North Carolina* presented a similar case of a struggle between two states over the waters of an interstate river.[31] Therefore, the sovereign states invoked the Court's original jurisdiction seeking an equitable apportionment of the river after failed attempts to negotiate.[32] The question before the Supreme Court of the United States was whether three nonsovereign, nonstate entities, the CRWSP, Duke Energy, and the city of Charlotte, had asserted compelling interests such that intervention was proper where those interests would be nominally represented by the sovereign states as parties to the original jurisdiction litigation.[33]

A. History of Original Jurisdiction

At the drafting of Article III of the Constitution, the Founders, realizing the importance of state autonomy and dignity, granted original jurisdiction to the Supreme Court "[i]n all Cases affecting Ambassadors, other public Ministers and Consuls, and those in which a State shall be Party, . . ."[34] Although this clause has always been described as self-executing, the

29. *See* Little, *supra* note 3, at 8, 9.

30. *See* Joseph W. Dellapenna, *Interstate Struggles Over Rivers: The Southeastern States and the Struggle Over the 'Hooch*, 12 N.Y.U. ENVTL. L.J. 828, 881–82 (2005) (citations omitted).

31. *See generally* Complaint, *supra* note 13.

32. *Id.* at 9–10.

33. *South Carolina*, 130 S. Ct. at 859.

34. U.S. CONST. art. III, § 2; *see* Vincent L. McKusick, *Discretionary Gatekeeping: The Supreme Court's Management of its Original Jurisdiction Docket Since 1961*, 45 ME. L. REV. 185, 186 (1993); *see also* Note, *The Original Jurisdiction of the United States Supreme Court*, 11 STAN. L. REV. 665, 665 (1959) (justifying the grant of original jurisdiction "to insure that issues involving basic adjustments within the federal structure or involving United States relationships with foreign countries would be heard before a tribunal whose prestige was commensurate with that of the parties before it.").

Founders further legislated this grant of jurisdiction at the First Congress in the Judiciary Act of 1789, giving the Supreme Court exclusive jurisdiction over suits between two or more states.[35] The need for such a clause is evident, as the states required a neutral and dignified forum in which to settle disputes among themselves that would "have been resolved by war or diplomatic negotiations prior to the formation of the federal union."[36] By accepting the Constitution, the states are deemed to have consented to this exercise of exclusive jurisdiction that comprises the vast majority of the original docket.[37]

Although the Court has the power to hear these cases under its exclusive, original jurisdiction, that power is discretionary and seldom exercised.[38] In 1900, the Court described this exercise of jurisdiction as "of so delicate and grave a character that it was not contemplated that it would be exercised save when the necessity was absolute and the matter in itself properly justiciable."[39] Over time, the Court's reluctance to exercise original jurisdiction has further increased, prompting Justice Harlan to state that "changes in the American legal system and the development of American society have rendered untenable, as a practical matter, the view that this Court must stand willing to adjudicate all or most legal disputes that may arise" under its original jurisdiction.[40] In 1992, Chief Justice Rehnquist emphasized that the exercise of original jurisdiction is to be used sparingly, and that "'[t]he model case for invocation of this Court's original jurisdiction is a dispute between States of such seriousness that it would amount to *casus belli*[41] if the States were fully sovereign.'"[42]

---

35. *See* 28 U.S.C. § 1251(a); McKusick, *supra* note 34, at 187; *see also* 20 CHARLES ALAN WRIGHT & MARY KAY KANE, FEDERAL PRACTICE AND PROCEDURE § 116 (2002). Although there was some confusion initially as to whether all original jurisdiction cases fell within the exclusive jurisdiction of the Court, it is now settled law that the Court's jurisdiction is only exclusive in actions between two or more states. *Id.*

36. Note, *supra* note 34, at 669. The Supreme Court has analogized "between the settlement of international and interstate disputes in determining the extent of its power to resolve controversies between the states." *Id.* (citations omitted).

37. WRIGHT & KANE, *supra* note 35, at 1092 (citing New Jersey v. New York, 30 U.S. 284 (1831)).

38. In these original jurisdiction cases, the Court will apply "federal common law." Hinderlider v. La Plata River & Cherry Creek Ditch Co., 304 U.S. 92, 110 (1938). *See* McKusick, *supra* note 34, at 187–88 (detailing the limited use of original jurisdiction) (citations ommitted).

39. Louisiana v. Texas, 176 U.S. 1, 15 (1900).

40. Ohio v. Wyandotte Chems. Corp., 401 U.S. 493, 497 (1971).

41. *Casus belli* is Latin for "[a]n act or circumstance that provokes or justifies war." BLACK'S LAW DICTIONARY 247 (9th ed. 2009).

42. Mississippi v. Louisiana, 506 U.S. 73, 76, 77 (1992) (quoting Texas v. New Mexico, 462 U.S. 554, 571 n.18 (1983)).

In a number of cases, states have been allowed to sue as *parens patriae*[43] to protect the interests of the states' populations as a whole.[44] In *Wyoming v. Colorado*, Wyoming brought suit as *parens patriae* to enjoin Colorado's diversion of the Laramie River that runs through both states.[45] This doctrine may also be used as a barrier to bringing suit under the Court's original jurisdiction where a party cannot show an interest beyond the protection of private rights to preserve Eleventh Amendment sovereign immunity.[46] In *North Dakota v. Minnesota*, the Court noted this distinction between a state's right to bring suit as *parens patriae* and "its lost power as a sovereign to present and enforce individual claims of its citizens as their trustee against a sister state."[47] North Dakota sued, invoking the Court's original jurisdiction and seeking damages on behalf of its citizens whose crops were destroyed when Minnesota straightened a river causing another North Dakota river to overflow.[48] The Court dismissed the action, characterizing it as a class action by injured North Dakota property owners barred by the Eleventh Amendment.[49] *Parens patriae* actions have also played a role in determining the adequacy of representation for purposes of considering whether intervention is proper.

B. Development of Nonsovereign Intervention

In original jurisdiction actions, the Supreme Court may use its discretion to allow the intervention of necessary parties.[50] Generally, *parens patriae* prevents individuals, private corporations, and municipalities from intervening in an exclusive jurisdiction action between states.[51] However, several cases in past decades have provided exceptions and circumstances to overcome this limitation.

---

43. *Parens patriae* means "parent of his or her country." Kathy Black, Comment, *Trashing the Presumption: Intervention on the Side of the Government*, 39 ENVTL. L. 481, 488 (2009). Derived from the English common law, states have used the doctrine to assert "sovereign interest in the natural resources within their borders." *Id.* at 489.

44. WRIGHT & KANE, *supra* note 35, at 1097.

45. Wyoming v. Colorado, 259 U.S. 419, 455 (1922).

46. U.S. CONST. amend. XI; WRIGHT & KANE, *supra* note 35, at 1094 (citations omitted). *See generally* Kentucky v. Indiana, 281 U.S. 163 (1930).

47. North Dakota v. Minnesota, 263 U.S. 365, 375–76 (1923).

48. *Id.* at 371–72.

49. *Id.* at 374–75.

50. FED. R. CIV. P. 24. The Supreme Court considers the Federal Rules of Civil Procedure a "guide" in original jurisdiction cases. Arizona v. California, 460 U.S. 605, 614 (1983) (citing Utah v. United States, 394 U.S. 89, 95 (1969)).

51. 32 AM. JUR. 2D *Federal Courts* § 514 (2007); 36 C.J.S. *Federal Courts* § 276 (2003). Under *parens patriae*, "the state is deemed to represent all its citizens in an original action and its view of the issues is deemed conclusive." 32 AM. JUR. 2D *Federal Courts* § 514 (2007) (citations omitted).

In *Oklahoma v. Texas*, the Supreme Court allowed intervention of private parties in an original jurisdiction action between two states.[52] Oklahoma sued Texas to settle a boundary dispute along the Red River to determine title to the southern half of the riverbed.[53] The United States intervened as a party shortly after Texas filed its answer in the matter, disputing the states' claims.[54] Upon a motion by the United States, approved by both states, the Court appointed a receiver to take possession of the disputed land, controlling the oil and gas operations.[55] The order also "provided for such interventions in the suit as would permit all possible claims to the property and proceeds in the receiver's possession to be freely and appropriately asserted."[56] Several parties then intervened to assert rights to the land held by the receiver that conflicted with the claims of the sovereign parties.[57] The Court determined that it would hear the intervenors' claims because no other court could lawfully interfere with the receiver's possession or control.[58] The Court did not provide a definite rule for nonsovereign intervention until 1953 when it decided *New Jersey v. New York*.[59]

In *New Jersey*, the Court held that "[a]n intervenor whose state is already a party should have the burden of showing some compelling interest in his own right, apart from his interest in a class with all other citizens and creatures of the state, which interest is not properly represented by the state."[60] In 1929, New Jersey brought an action against New York state and New York City under the Court's original jurisdiction to enjoin New York's proposed diversion of the Delaware River.[61] The Court then granted a motion to intervene in the action *pro interesse suo* immediately filed by Pennsylvania.[62] In 1931, the Court entered a decree enjoining both

---

52. *See generally* Oklahoma v. Texas, 258 U.S. 574 (1922).

53. *Id.* at 578. Oklahoma, Texas, the United States, and several private individuals and entities were fighting over title to the property because the southerly half of the riverbed had been "recently discovered to be underlaid with strata bearing oil and gas and to be of great value by reason thereof." *Id.* at 579. The court described a mad rush to drill the land that resulted in armed conflicts and the calling of the Texas state militia. *Id.* at 579–80.

54. *Id.* at 578–79.

55. *Id.* at 580.

56. *Id.* (citing Oklahoma v. Texas, 252 U.S. 372 (1920)).

57. *Id.* at 581.

58. *Id.* The Court went on to say that "[i]t long has been settled that claims to property or funds of which a court has taken possession and control through a receiver or like officer may be dealt with as ancillary to the suit wherein the possession is taken and the control exercised . . . ." *Id.* (citations omitted).

59. *See generally* New Jersey v. New York, 345 U.S. 369 (1953) (per curiam).

60. *Id.* at 373.

61. *Id.* at 370. New Jersey joined New York City as a defendant because the river diversions were actually being planned by and for the city, acting under the state's authority. *Id.* at 370–71.

62. *Id.* at 371 (citation omitted).

New York state and New York City from diverting more than 440 mgd from the river.[63] In 1952, the Court granted New York City's motion to modify the decree.[64] Later in the year, the city of Philadelphia sought leave to intervene, asserting its interest in the river.[65] All of the parties to the litigation opposed the motion, arguing that the intervention would allow a suit against a state by a citizen of another state barred by the Eleventh Amendment, that Pennsylvania represents the interests of Philadelphia as *parens patriae*, and "that intervention should be denied, in any event, as a matter of sound discretion."[66]

In determining whether to grant Philadelphia's motion to intervene, the Court described the doctrine of *parens patriae* as "a necessary recognition of sovereign dignity, as well as a working rule for good judicial administration."[67] Under the rule, the Court denied Philadelphia's motion to intervene because the city was properly represented by Pennsylvania.[68] Further, Philadelphia failed to meet its burden of proving a compelling interest apart from the interests of other Pennsylvania citizens.[69] This rule requiring a compelling interest, first articulated by the Court in *New Jersey*, remains the standard today in original jurisdiction cases.[70] The Court had opportunities to determine sufficient compelling interests in *Texas v. Louisiana* in 1976, *Maryland v. Louisiana* in 1981, and *Arizona v. California* in 1983.[71]

In *Texas v. Louisiana*, the Court allowed the city of Port Arthur, Texas, to intervene to protect its interests in the island claims of the United States.[72] Later, the Court permitted seventeen pipeline companies to

---

63. *Id.* (citations omitted).

64. *Id.* (citation omitted).

65. *Id.* at 372.

66. *Id.*

67. *Id.* at 373. The Court also observed that without the doctrine, "a state might be judicially impeached on matters of policy by its own subjects, and there would be no practical limitation on the number of citizens, as such, who would be entitled to be made parties." *Id.*

68. *Id.* The Court feared an opening of the floodgates of additional parties seeking to intervene, and warned that "[o]ur original jurisdiction should not be thus expanded to the dimensions of ordinary class actions." *Id.*

69. *Id.* at 373–74.

70. *See* Nebraska v. Wyoming, 515 U.S. 1, 21–22 (1995) (appeal following Nebraska's attempt to modify a 1945 decree equitably apportioning the North Platte River); United States v. Nevada, 412 U.S. 534, 538 (1973) (per curiam) (where United States sought a declaration of rights to the Truckee River that flows through Nevada and Colorado); Illinois v. Milwaukee, 406 U.S. 91, 97 (1972) (where Illinois sought leave to file a complaint against four Wisconsin cities and two local sewerage commissions for polluting Lake Michigan).

71. *See generally* Arizona v. California, 460 U.S. 605 (1983); Maryland v. Louisiana, 451 U.S. 725 (1981); Texas v. Louisiana, 426 U.S. 465 (1976) (per curiam).

72. *Texas*, 426 U.S. at 466 (citation omitted).

intervene as plaintiffs in an original action in *Maryland v. Louisiana*.[73] In that case, several states, the United States, and the nonsovereign pipeline companies challenged the constitutionality of a "First-Use Tax" enacted in Louisiana that affected certain uses of natural gas brought into Louisiana.[74] The Court determined that intervention was proper because the pipeline companies, as owners of the gas subject to the tax, had a direct stake in the litigation.[75] Two years later, the Court allowed several Indian tribes to intervene to protect their rights to the Colorado River in *Arizona v. California*.[76]

In 1952, Arizona filed a motion for leave to file a complaint against California and several California public agencies to equitably apportion the Colorado River as to the two states.[77] Shortly after the Court granted leave, the United States, on behalf of several Indian tribes, and Nevada intervened in the original jurisdiction action.[78] In 1964, the Court entered a decree in the case that was to be modified in 1979 before the Indian tribes, previously represented by the United States, sought to intervene in the litigation in 1977.[79] After initial opposition from both the states and the United States, the United States withdrew its opposition to the Indian tribes' intervention.[80] The Court, agreeing with the Special Master appointed in the original action, determined that the Indian tribes' motions to intervene should be granted because it would not contradict the Eleventh Amendment and the United States could not adequately represent the tribes' interests.[81] The Eleventh Amendment did not bar the suit because the tribes did not "seek to bring new claims or issues against the states, but only ask[ed] leave to participate in an adjudication of their vital water rights . . . ."[82] Moreover, the Court stated that "it is obvious that the Indian Tribes, at a minimum, satisfy the standards for permissive intervention set forth in the Federal Rules."[83] Because the Tribes' interests were subject to past and future litigation, the tribes have become recognized as sovereign entities as

73. *Maryland*, 451 U.S. at 745 n.21 (citations omitted).

74. *Id.* at 728.

75. *Id.* at 745 n.21 (citations omitted). The Court also determined that the companies' interventions would facilitate "a full exposition of the issues." *Id.* (citations omitted).

76. *See generally Arizona*, 460 U.S. at 615 (citations omitted).

77. *Id.* at 608.

78. *Id.* Utah and New Mexico were joined as defendants in the action. *Id.* The United States represented the interests of the Colorado River Indian Tribes, Fort Mojave Indian Tribe, Chemehuevi Indian Tribe, Cocopah Indian Tribe, and Fort Yuma (Quechan) Indian Tribe. *Id.* at 608–09.

79. *Id.* at 612.

80. *Id.*

81. *Arizona*, 460 U.S. at 613–14.

82. *Id.* at 614. The Court further stated that "our judicial power over the controversy is not enlarged by granting leave to intervene," and, accordingly, the Eleventh Amendment is not implicated. *Id.* (citing Maryland v. Louisiana, 41 U.S. 725, 745 (1981)).

83. *Id.* at 614–15; *see also* FED. R. CIV. P. 24.

"independent qualified members of the modern body politic,"[84] and the state parties could not show prejudice or undue delay, the Court granted the Indian tribes' motions to intervene.[85]

III. THE CRWSP AND DUKE ENERGY HAVE SHOWN COMPELLING INTERESTS TO JUSTIFY INTERVENTION IN THE ORIGINAL ACTION BETWEEN SOUTH CAROLINA AND NORTH CAROLINA

In *South Carolina v. North Carolina*, the United States Supreme Court, in a 5–4 decision, held that two nonsovereign entities, the CRWSP and Duke Energy, met their burden of proving a sufficiently compelling interest, apart from the interests of the other state's citizens and inadequately represented by the states, which justified intervention in the litigation before the Court under its original jurisdiction.[86] However, the Court also held that Charlotte did not meet its burden of proving a sufficiently compelling interest and that it was adequately represented by North Carolina, denying intervention.[87]

The Court began the opinion with an analysis of original jurisdiction actions involving nonstate entities and concluded that it is "not a novel proposition to accord party status to a citizen in an original action between States."[88] The Court also observed that it has, on several occasions in the past ninety years, granted leave to nonstate parties to intervene in original actions.[89] The Court next addressed the Special Master's rule, rejecting it as too broad by "account[ing] for the full compass of [Supreme Court] precedents."[90] Instead, it adopted the appropriate standard set forth in *New Jersey v. New York* which states that "[a]n intervenor whose state is already a party should have the burden of showing some compelling interest in his own right, apart from his interest in a class with all other citizens and creatures of the state, which interest is not properly represented by the state."[91]

In embracing the *New Jersey* standard, the Court acknowledged the difficulties of meeting such a high burden, but justified it by recognizing

---

84. *Arizona*, 460 U.S. at 615 (quoting Poafpybitty v. Skelly Oil Co., 390 U.S. 365, 369 (1968)).

85. *Id.*

86. South Carolina v. North Carolina, 130 S. Ct. 854, 859 (2010).

87. *Id.*

88. *Id.* at 861–62.

89. *Id.* at 862. In discussing *Oklahoma v. Texas*, Justice Alito briefly addressed the minority in stating that the case "counsel[s] against inferring from our precedents, as THE CHIEF JUSTICE does with respect to equitable apportionment actions, a rule against non-state intervention in such 'weighty controversies.'" *Id.* at 862 n.3 (emphasis in original) (citing Oklahoma v. Texas, 254 U.S. 609 (1920)).

90. *Id.* at 862. Because it declined to adopt the Special Master's rule, the Court did not address South Carolina's exception to the proposed rule. *Id.* at 862 n.4.

91. *Id.* at 862, 863 (quoting New Jersey v. New York, 345 U.S. 369, 373 (1953)).

that original actions "tax the limited resources of this Court by requiring us 'awkwardly to play the role of factfinder' and diverting our attention from our 'primary responsibility as an appellate tribunal.'"[92] The Court concluded that, although the threshold for intervention is high, it is not "insurmountable."[93] The Court then addressed the three motions for intervention from the nonsovereign entities, applying the *New Jersey* standard and beginning with the CRWSP.[94]

The Court determined that the CRWSP should be allowed to intervene because it demonstrated a sufficiently compelling interest, and neither North Carolina nor South Carolina could adequately represent that interest.[95] It first analyzed the peculiar circumstances of the bistate entity, noting that "it is an unusual municipal entity, established as a joint venture with the encouragement of regulatory authorities in both States and designed to serve the increasing water needs of Union County, North Carolina and Lancaster County, South Carolina."[96] Furthermore, the entity draws and transfers water from the Catawba River with the authority of both states because the water is drawn from below the Lake Wylie dam in South Carolina, and pumped across the state border by virtue of a parallel certificate issued by North Carolina.[97]

In determining that the CRWSP's interests were distinct from those of other citizens of the states, the Court emphasized the $30 million investment that the CRWSP made in its plant that Union and Lancaster Counties had incurred as debt.[98] The Court determined that "[a]ny disruption to the CRWSP's operations would increase—not lessen—the difficulty of our task in achieving a 'just and equitable' allocation in this dispute."[99] The Court then addressed whether South Carolina or North Carolina could adequately represent the CRWSP's interests and determined that neither state could do so.[100]

Because both states will argue that the other state's equitable share should be reduced, the Court concluded that denying intervention to the CRWSP would upset the fine balance on which the joint venture is based

92. *Id.* (quoting Ohio v. Wyandotte Chems. Corp., 401 U.S. 493, 498 (1971); Maryland v. Louisiana, 451 U.S. 725, 762 (1981) (Rehnquist, J., dissenting)).

93. *Id.* at 864.

94. *Id.*

95. *Id.*

96. *Id.* Justice Alito also noted that the CRWSP "has an advisory board consisting of representatives from both counties, draws its revenues from its bistate sales, and operates infrastructure and assets that are owned by both counties as tenants-in-common." *Id.*

97. *Id.* at 864, 865.

98. *Id.* at 865. The counties are equally responsible for the operating costs of the CRWSP, which is "designed to break even from year to year." *Id.*

99. *Id.*

100. *Id.*

because "neither State has sufficient interest in maintaining that balance to represent the full scope of the CRWSP's interests."[101]

After overruling South Carolina's exception and allowing the CRWSP to intervene, the Court turned to Duke Energy's motion to intervene. The Court began its analysis with an examination of the process by which interstate streams are justly and equitably apportioned to place Duke Energy's interests in context.[102] The Court described the process in terms of a multiple factor analysis[103] with the "'exercise of an informed judgment'"[104] and without "hesitat[ion] to seek out the most relevant information from the source best situated to provide it."[105]

The Court then turned to its examination of Duke Energy's interests. Because Duke Energy provides electricity for the region using the waters of the Catawba River,[106] the Court determined that any equitable apportionment of the river would have to account for the allocation that the entity would require to maintain its operations and to continue to power the region.[107] The Court emphasized "the appropriateness of considering 'the balance of harm and benefit that might result' from a State's proposed diversion of a river" that would give Duke Energy a powerful interest in shaping the outcome of the litigation.[108] Furthermore, because there is no other similarly situated entity on the Catawba River, Duke Energy's interests are set "apart from the class of all other citizens of the States."[109] The Court also stressed the importance of Duke Energy's interest "in protecting the terms of its existing FERC license and the CRA[110] that forms the basis of Duke Energy's pending renewal application."[111] As the CRA "represents the full consensus of 70 parties from both States" regarding

---

101. *Id.* The Court also mentioned the "further complication" that South Carolina may not sufficiently protect all the uses of Lancaster County's apportioned share because that county "has an obligation to provide water service to certain customers in Mecklenburg County, North Carolina." *Id.* at 865 n.6.

102. *Id.* at 866.

103. The Court set forth several relevant factors considered in the apportionment process including the "'physical and climatic conditions, the consumptive use of water in the several sections of the river, the character and rate of return flows, the extent of established uses, the availability of storage water, [and] the practical effect of wasteful uses on downstream areas, . . .'" *Id.* at 866 (quoting Colorado v. New Mexico, 459 U.S. 176, 183 (1982)).

104. *Id.* at 866 (quoting *Colorado*, 459 U.S. at 183).

105. *Id.* at 866 (quoting Maryland v. Louisiana, 451 U.S. 725, 745 n.21 (1981)).

106. *See supra* note 15 and accompanying text (describing Duke Energy's hydroelectric operations on the river).

107. *South Carolina*, 130 S. Ct. at 866 (citing *Colorado*, 459 U.S. at 188).

108. *Id.* at 866 (quoting *Colorado*, 459 U.S. at 188).

109. *Id.* at 866 (citing New Jersey v. New York, 345 U.S. 369, 373 (U.S. 1953)).

110. *See supra* note 15 and accompanying text (explaining Duke Energy's FERC license and the CRA).

111. *South Carolina*, 130 S. Ct. at 866.

relevant factors[112] that the Court will likely consider in reaching its ultimate disposition of the case, the Court determined that Duke Energy met the high burden of showing "unique and compelling interests" to justify intervention.[113]

However, the Court then evaluated whether Duke Energy's interests could be adequately represented by one of the sovereign entities in the original action, but concluded that neither State is properly situated to do so.[114] The Court reasoned that because neither North Carolina nor South Carolina had signed the CRA or expressed an intention to defend its terms,[115] Duke Energy's vital and relevant "interests should be represented by a party [to the] action."[116] The Court overruled South Carolina's exception to the intervention and granted Duke Energy's motion to intervene in the original action.[117]

The Court then began its evaluation of the third and final motion to intervene filed by Charlotte, North Carolina.[118] In its examination of Charlotte's interests in the original action, the Court focused on South Carolina's complaint, noting that although the municipality was named as an entity that transfers large amounts of water from the river, the complaint did not seek relief against Charlotte directly but "against all North Carolina-authorized transfers of water from the Catawba River basin, 'past or future,' in excess of North Carolina's equitable share."[119] Based on this information, the Court determined Charlotte to be a member of the class of North Carolina water consumers, and found its large transfers distinguished it only in degree from the other members in the class.[120] The Court also distinguished Charlotte's interests from the CRWSP because Charlotte's interests did not "fall on both sides of this dispute, . . . such that the viability of Charlotte's operations . . . [were] called into question."[121]

After determining that Charlotte's interests were not sufficiently compelling, the Court then determined that North Carolina could

---

112. *Id.* at 866–67. These factors included "the appropriate minimum continuous flow of Catawba River water into South Carolina under a variety of natural conditions and, in times of drought, the conservation measures to be taken by entities that withdraw water from the Catawba River." *Id.* at 867.

113. *Id.* at 867.

114. *Id.*

115. *Id.* (citation omitted). In fact, North Carolina "expressed an intention to seek its modification." *Id.*

116. *Id.*

117. *Id.*

118. *Id.* The Court noted that "Charlotte is a municipality of North Carolina, and for purposes of this litigation, its transfers of water from the Catawba River basin constitute part of North Carolina's equitable share." *Id.*

119. *Id.* (quoting Complaint at 10, *South Carolina*, 130 S. Ct. 854 (2010) (No. 138, Original)).

120. *Id.* (citing New Jersey v. New York, 345 U.S. 369, 373 (U.S. 1953)).

121. *Id.*

adequately represent Charlotte as *parens patriae*.[122] The Court, citing *New Jersey v. New York*, acknowledged that "a State's sovereign interest in ensuring an equitable share of an interstate river's waters is precisely the type of interest that the State, as *parens patrie* [sic], represents on behalf of its citizens."[123] Thus, absent showing a "compelling interest . . . distinct from the collective interest of 'all other citizens and creatures of the state,'" the Court must conclude that "North Carolina properly represents Charlotte in this dispute over a matter of uniquely sovereign interest."[124] Therefore, deciding that Charlotte's interest is not sufficiently unique and that it will be properly represented by North Carolina as *parens patriae*, the Court sustained South Carolina's exception and denied Charlotte's motion to intervene.[125] The Court concluded by overruling South Carolina's exceptions to the Special Master's report as to the CRWSP and Duke Energy, and sustaining South Carolina's exceptions to the Special Master's report as to Charlotte.[126]

In a separate opinion concurring in part and dissenting in part,[127] authored by Chief Justice Roberts and joined by Justices Thomas, Ginsburg, and Sotomayor, the dissent disagreed with the Court's holding that the CRWSP and Duke Energy should be permitted to intervene, but agreed that Charlotte should not be permitted to intervene.[128] After determining that the majority misapplied the *New Jersey* test, the dissent began its analysis by emphasizing that "[t]he result is literally unprecedented, . . . this Court has never before granted intervention in such a case to an entity other than a State, the United States, or an Indian tribe. Never."[129] The dissent further stated that equitable "apportionment of an interstate waterway is a sovereign dispute, and the key to intervention in such an action is just that—sovereignty."[130] In opposing the Court's decision, the dissent predicted that permitting nonsovereign intervention

---

122. *Id.*

123. *Id.* (citations omitted).

124. *Id.* at 867–68 (citations omitted). The Court also highlighted the importance of North Carolina's statements during the course of the proceedings to support its conclusion. *Id.* at 868. It noted, "North Carolina has said that it will defend Charlotte's authorized 33 mgd transfer." *Id.* (citation omitted).

125. *Id.* The Court also dismissed as inappropriate Rule 24 analysis for permissive intervention due to "North Carolina's adequate representation of Charlotte and the heightened standard for intervention in original actions." *Id.* at 868 n.8 (citing New Jersey v. New York, 345 U.S. 369, 373 (1953)); *see also* FED. R. CIV. P. 24(b).

126. *South Carolina*, 130 S. Ct. at 868.

127. Although concurring in part and dissenting in part, the author will refer to the minority as "the dissent," as the separate opinion focused primarily on the dissenting points of the CRWSP and Duke Energy's intervention.

128. *South Carolina*, 130 S. Ct. at 868–69 (Roberts, C.J., concurring in part and dissenting in part).

129. *Id.* at 869 (Roberts, C.J., dissenting).

130. *Id.* (Roberts, C.J., dissenting).

"has the potential to alter in a fundamental way the nature of our original jurisdiction, transforming it from a means of resolving high disputes between sovereigns into a forum for airing private interests."[131]

The dissent then discussed the two basic principles that have governed the Supreme Court's exercise of its original jurisdiction.[132] The first principle the dissent set forth was an understanding that the Court's original jurisdiction "was granted to provide a forum for the peaceful resolution of weighty controversies involving the States."[133] The dissent reiterated that original jurisdiction may be invoked for the resolution of state claims, not private claims.[134] The second pragmatic principle is that the Court "[is] not well suited to assume the role of a trial judge."[135] The dissent then concluded that these considerations are reflected in the *New Jersey* standard for intervention in original actions.[136] The application of the *New Jersey* rule, the dissent explained, "precludes a State from being 'judicially impeached on matters of policy by its own subjects,' and prevents the use of the Court's original jurisdiction to air 'intramural dispute[s]' that should be settled in a different forum—namely, within the States."[137] The Court then turned its attention to equitable apportionment actions in general.

In support of its observation that the Court has never allowed a nonsovereign entity to intervene in an equitable apportionment action, the dissent reasoned that "[a]n interest in water is an interest shared with other citizens, and is properly pressed or defended by the State" and deemed private entities' interests "intramural disputes" that are settled within the state after the water is apportioned between the states.[138] The dissent then reiterated that "[t]he interests of a State's citizens in the use of water derive entirely from the State's sovereign interest in the water way [and] [i]f the State has no claim to the waters of an interstate river, then its citizens have none either."[139] Therefore, the state "'must be deemed to represent *all* its citizens,' not just those who subscribe to the State's position before this

131. *Id.* (Roberts, C.J., dissenting).

132. *Id.* (Roberts, C.J., dissenting).

133. *Id.* (citing Louisiana v. Texas, 176 U.S. 1, 15 (1900)).

134. *Id.* (citations omitted).

135. *Id.* (citing Ohio v. Wyandotte Chems. Corp., 401 U.S. 493, 497 (1971)). The dissent then referenced the Court's use of Special Masters in original actions to cope with this suitability principle. *Id.*; *see supra* note 19 and accompanying text (explaining the Court's use of Special Masters in original jurisdiction actions).

136. *Id.* at 869–70.

137. *Id.* at 870 (quoting New Jersey v. New York, 345 U.S. 369, 373 (1953)) (alteration in original). The dissent also noted that the rule, when properly applied, "provides a much-needed limiting principle that prevents the expansion of our original proceedings 'to the dimensions of ordinary class actions,' or 'town-meeting lawsuits.'" *Id.* (quoting *New Jersey*, 345 U.S. at 376) (Jackson, J., dissenting)) (citation omitted).

138. *Id.* (quoting *New Jersey*, 345 U.S. at 373).

139. *Id.* (citing Hinderlider v. La Plata River & Cherry Creek Ditch Co., 304 U.S. 92, 102 (1938)).

Court."[140] The dissent added that individual citizens are not required to "be made parties to an equitable apportionment action because the Court's judgment in such an action does not determine the water rights of any individual citizen."[141] In examining the extensive history of rejecting attempts by nonstate entities to intervene in equitable apportionment actions, the dissent concluded, "private entities can rarely, if ever, intervene in original actions involving the apportionment of interstate waterways."[142]

Finally, the dissent addressed the cases relied upon by the majority and emphasized that none were equitable apportionment actions under the Court's original jurisdiction.[143] The dissent dismissed the cases as irrelevant because the intervenor interests in those cases were distinct from the general shared interest in water.[144] First, the dissent distinguished *Arizona v. California* where the Court allowed Indian tribes to intervene because they were *sovereign* entities.[145] Second, the dissent observed that *Texas v. Louisiana*[146] and *Oklahoma v. Texas*[147] were boundary dispute cases and concluded that "[a] claim to title in a particular piece of property is quite different from a general interest shared by all citizens in the State's waters."[148] Third, *Maryland v. Louisiana*,[149] the dissent concluded, involved "an interest in a tax imposed only on discrete parties [that] is obviously different from a general interest shared by all citizens of the State."[150] The dissent then turned to its analysis of the three motions to intervene filed by the nonstate entities.

Evaluating the interests of the three entities, the dissent emphasized that all of these different interests are based in a common, shared interest in the

---

140. *Id.* at 871 (quoting *New Jersey*, 345 U.S. at 372) (emphasis added).

141. *Id.*

142. *Id.* (citing United States v. Nevada, 412 U.S. 534, 538 (1973)). The dissent then addressed the majority's contention "that this dissent reads our precedents to establish 'a rule against nonstate intervention' in equitable apportionment actions." *Id.* at 872 n.1. *See supra* note 86 and accompanying text. Conceding that "[a] private party (or perhaps a Compact Clause entity) with a federal statutory right to a certain quantity of water might have a compelling interest in an equitable apportionment action that is not fairly represented by the States," the dissent denied any presumption against nonstate intervention. *South Carolina*, 130 S. Ct. at 872 n.1 (Roberts, C.J., dissenting).

143. *Id.* at 872.

144. *Id.*

145. *Id.* (emphasis added) (citation omitted); *see* Arizona v. California, 460 U.S. 605 (1983).

146. Texas v. Louisiana, 426 U.S. 465 (1976).

147. Oklahoma v. Texas, 258 U.S. 574 (1922). The dissent also noted in this case that the individual citizens' intervention was permitted only because the Court took receivership of the land and no other court could disrupt that possession. *South Carolina*, 130 S. Ct. at 872 (Roberts, C.J., dissenting).

148. *South Carolina*, 130 S. Ct. at 872 (Roberts, C.J., dissenting).

149. Maryland v. Louisiana, 451 U.S. 725 (1981).

150. *South Carolina*, 130 S. Ct. at 873 (Roberts, C.J., dissenting).

Catawba River's waters.[151] Although every entity may claim a different use for the water, their interests in the water are not unique based on that fact alone.[152] In evaluating the Court's decision to grant Duke Energy's motion to intervene, the dissent disagreed that Duke Energy's possession of valuable and relevant information was sufficient to justify intervention.[153] The dissent strongly suggested that Duke Energy could better provide its information by participating in the litigation as *amicus curiae*, and not as a party to the original action.[154] Finally, the dissent determined that Duke Energy's interests in providing power to the region and protecting its FERC license were merely "an articulation of the *reason* Duke Energy asserts a particular interest," and that "[w]eighing those interests is an 'intramural' matter for the State."[155]

Turning to the CRWSP, the dissent determined that the CRWSP's interests were not distinct from Charlotte's interests and warned of allowing intervention based on the CRWSP's bistate nature alone.[156] The dissent explained that allowing an exception to the *New Jersey* standard, based merely on an "intermingling of state interests," would render the standard useless because it would allow intervention "of any bistate entity, or indeed any corporation or individual conducting business in both States."[157] In regard to both Duke Energy and the CRWSP, the dissent relied on the principle of *parens patriae* in disagreeing that the two entities will not be adequately represented by the states.[158] The dissent pointed out that the majority overlooked the Court's decision in *New Jersey* and the doctrine of *parens patriae* by determining that the entities would not be properly represented merely because the States may adopt positions contrary to the entities.[159]

In opposing the majority's grants of the CRWSP and Duke Energy's motions to intervene, the dissent presaged "[a]n equitable apportionment action will take on the characteristics of an interpleader case, with all those asserting interests in the limited supply of water jostling for their share like animals at a watering hole."[160] The dissent explained that permitting the nonsovereign entities to intervene would certainly protract the resolution of

---

151. *Id.* (Roberts, C.J., dissenting).

152. *Id.* (Roberts, C.J., dissenting).

153. *Id.* (Roberts, C.J., dissenting).

154. *Id.* (Roberts, C.J., dissenting).

155. *Id.* at 873–74 (quoting New Jersey v. New York, 345 U.S. 369, 373 (1953)) (emphasis in original). The dissent also noted that "the Federal Government is doubtless familiar with the pending FERC proceedings, and it sees no corresponding need for us to grant Duke Energy's motion to intervene." *Id.* at 874 (citation omitted).

156. *Id.*

157. *Id.* (quoting *South Carolina*, 130 S. Ct. at 865 n.6 (majority opinion)).

158. *Id.* (Roberts, C.J., dissenting).

159. *Id.* (Roberts, C.J., dissenting).

160. *Id.* at 875.

what is already a time-intensive matter with "more issues to decide, more discovery requests, [and] more exceptions to the recommendations of the Special Master."[161] In its conclusion, the dissent reiterated its proposal that the CRWSP and Duke Energy participate as *amici curiae* instead of intervening parties to ensure the effective and timely disposition of the already lengthy equitable apportionment action.[162] Chief Justice Roberts closed stating "I would grant South Carolina's exceptions, and deny the motions to intervene," determining that "a squabble among private entities within a State over how to divvy up that State's share [of water] does not" present a case for invoking the Court's original jurisdiction.[163]

IV. CONSEQUENCES OF SOUTH CAROLINA V. NORTH CAROLINA

The Supreme Court's decision in *South Carolina v. North Carolina* evinces apparent uncertainty regarding the *parens patriae* doctrine coupled with an expansionary ruling in the Court's struggle to manage its original jurisdiction docket.[164] Moreover, the very unusual 5–4 split in the *South Carolina* Court illustrates the disagreement in determining how to manage and whether to expand the Court's original jurisdiction docket.[165]

Although applying the same *New Jersey* standard in ruling on the motions to intervene, the majority and the dissent arrived at divergent decisions based on each side's determination of the scope of the ruling.[166] The majority adopted a broad scope couched in terms of original jurisdiction in general.[167] Although it refused to adopt the Special Master's rule because of its breadth of application in original jurisdiction actions, the Court relied upon the same precedent to allow intervention in this equitable apportionment action.[168] The scope of the Court's decision encompassed

---

161. *Id.* The Court also observed that "intervention makes settling a case more difficult, as a private intervenor has the right to object to a settlement agreement between the States, if not the power to block a settlement altogether." *Id.* (citation omitted).

162. *Id.*

163. *Id.* at 876.

164. *See* McKusick, *supra* note 34, at 189–90 ("[T]he Supreme Court often rejects original jurisdiction cases in summary orders that give no suggestion of either the subject matter of the attempted suits or the Court's reasons for rejecting them.").

165. As will be discussed later in detail, the Court did not split along traditional political lines in *South Carolina*. Traditionally conservative and liberal justices fell on both sides of the decision.

166. *South Carolina*, 130 S. Ct. at 869 (Roberts, C.J., dissenting).The dissent argued that the majority incorrectly applied the *New Jersey* standard. *Id.*

167. *See id.* at 861 (majority opinion).

168. *Id.* at 862. *See generally* Arizona v. California, 460 U.S. 605 (1983) (equitable apportionment); Maryland v. Louisiana, 451 U.S. 725 (1981) (constitutional challenge of natural gas tax); Texas v. Louisiana, 426 U.S. 465 (1976) (boundary dispute); New Jersey v. New York, 345 U.S. 369 (1953) (equitable apportionment); Oklahoma v. Texas, 258 U.S. 574 (1922) (boundary dispute).

original jurisdiction actions in general, focusing on *South Carolina* as an original jurisdiction action rather than an equitable apportionment action, accounting for original jurisdiction precedent, and permitting the nonsovereign entities to intervene.[169]

The dissent, however, narrowed the scope of its decision couched in terms of equitable apportionment actions within the Court's original jurisdiction.[170] The dissent chided the majority for its reliance on cases outside the scope of equitable apportionment actions in allowing the nonsovereign entities to intervene,[171] limiting the scope of its analysis to equitable apportionment actions invoking the Court's original jurisdiction.[172] Ultimately, the narrow scope set forth in the dissent is more appropriate considering the limitations of the Supreme Court as a court of first impression in original jurisdiction actions.[173]

After setting the scope of its analysis, the majority then proceeded to misapply the *New Jersey* standard and overlook the doctrine of *parens patriae* almost entirely by distinguishing compelling interests different in degree, but not in kind from the other citizens of the states.[174] As the dissent pointed out, the Court incorrectly determined that the CRWSP and Duke Energy demonstrated compelling interests distinct from their interests in a class with all other citizens of the state.[175] Although Duke Energy has significant interests in the Catawba River as an energy producer and provider, this interest is only different in degree from the other citizens of the state as consumers.[176] Similarly, the Court overlooked the impact of the CRWSP's intervention and justified its intervention based on its bistate interests alone.[177] The dissent commented that under the majority's reasoning "any bistate entity, or indeed any corporation or individual conducting business" between states would then be allowed to intervene merely by virtue of its bistate operations, ignoring the doctrine of *parens patriae*.[178]

Under *parens patriae*, a state is deemed to represent the interests of all of its citizens regardless of whether they are contrary to the state's interests.[179] It seems that the *South Carolina* Court determined that *parens patriae* would only apply in Charlotte's case as a municipality of North

169. *South Carolina*, 130 S. Ct. at 862 (majority opinion).

170. *Id.* at 870 (Roberts, C.J., dissenting).

171. *See generally id.* at 870–71.

172. *Id.* at 873.

173. *See supra* notes 38–41 and accompanying text; *see also supra* note 135 and accompanying text.

174. *See generally South Carolina*, 130 S. Ct. at 874–75 (Roberts, C.J., dissenting).

175. *Id.* at 873.

176. *Id.* at 874.

177. *Id.* at 865 (majority opinion).

178. *Id.* at 874 (Roberts, C.J., dissenting).

179. *See generally id.* at 873.

Carolina whose "interest falls squarely within the category of interests with respect to which a State must be deemed to represent all of its citizens."[180] The Court then went on to say that "respect for 'sovereign dignity' requires us to recognize that North Carolina properly represents Charlotte in this dispute over a matter of uniquely sovereign interest."[181]

Following contradictory reasoning, the Court determined the equitable apportionment of water to be a "uniquely sovereign interest" that the state protects as *parens patriae*, but then ignored the doctrine in determining that the CRWSP and Duke Energy would not be adequately represented by the states.[182] The Court contended that the CRWSP will not be adequately represented because each state will argue that the other should receive less water, upsetting the balance on which the joint venture is based.[183] As the dissent suggested, the Court determined that the states could not adequately represent the CRWSP because they will advance positions contrary to the CRWSP's interest in direct contravention of *parens patriae*.[184] Likewise, the Court ignored the *parens patriae* doctrine again in determining that Duke Energy would be inadequately represented merely because neither state had committed to signing or defending the CRA.[185] Ultimately, the Court reached conflicting results by disregarding the *parens patriae* doctrine for the corporate entities, vaguely attempting to distinguish the CRWSP and Duke Energy.[186] Thus, the Court contradicted its ruling that "due respect [be] given to 'sovereign dignity.'"[187]

This contradictory decision cannot be explained by the political affiliations that so often seem to guide Supreme Court decisions. The majority opinion, written by conservative Justice Alito, was joined by two traditionally liberal Justices, Breyer and Stevens, moderate Justice Kennedy, and conservative Justice Scalia.[188] The dissenting opinion, written by conservative Chief Justice Roberts, was joined by liberal Justices Ginsburg and Sotomayor, and conservative Justice Thomas.[189] With Justices of both ideological orientations taking places on each side of the Court's decision, there is no clear indication of politics in *South Carolina*. The split could be attributed to the majority's desire to protect the business interests of the corporations in the action, but not the municipality.[190] However, this desire to protect corporate business interests is regarded as a

180. *Id.* at 867 (majority opinion).
181. *Id.* at 868.
182. *See generally id.* at 864–67.
183. *Id.* at 865.
184. *Id.* at 874 (Roberts, C.J., dissenting).
185. *Id.* at 867 (majority opinion).
186. *Id.* at 863 (quoting New Jersey v. New York, 345 U.S. 369, 373 (1953)).
187. *Id.* (quoting *New Jersey*, 345 U.S. at 373).
188. *Id.* at 858.
189. *Id.* at 868 (Roberts, C.J., dissenting).
190. *See generally id.* at 864–68 (majority opinion).

typically conservative interest. More likely, the Court split based on differing beliefs regarding how to manage the original jurisdiction docket.

Although the majority discussed the importance of exercising original jurisdiction sparingly,[191] this concern did not seem to play a significant role in the Court's decision to allow the intervention of the two nonsovereign entities. In fact, the Court never mentioned any reason why allowing the interventions would benefit the resolution of the case or whether the interventions would cause undue prejudice and delay to the state parties. The dissent, however, discusses the adverse effects of allowing the interventions, notably mentioning that the private intervenors now have "the right to object to a settlement agreement between the States, if not the power to block a settlement altogether."[192] Therefore, it appears that the majority favors an expansion of the Court's original jurisdiction while the dissent prefers to maintain the importance of the Court's primary role as a court of last resort.[193]

The Court's decision in *South Carolina* will have the practical effect of expanding the Court's original jurisdiction to include nonsovereign entities, taxing the Court's limited resources as a court of first impression, and leaving the Court in a difficult position to determine intervention on a case-by-case, entity-specific basis. Furthermore, the increasing demand for water engendered by economic and population growth will present the Court with more and more cases of equitable apportionment of the nation's interstate waterways arising under the Court's original jurisdiction, particularly if water becomes scarcer because of fundamental changes in the environment. After *South Carolina*, any corporate entity asserting a commercial interest in the waters of an interstate stream will likely be permitted to intervene to affect the outcome of the equitable apportionment. Beyond enlarging the Court's duties to settle matters between and within the states, the *South Carolina* decision will, unfortunately, detract from the Court's role as an appellate court by increasing the original jurisdiction docket. Thus, practitioners looking to intervene in an original action in order to shape the outcome for its corporate clients need only prove that their clients have significant business interests subject to the litigation that cannot be represented by the states because such interest is adverse to the states' interest.

---

191. *Id.* at 863.

192. *Id.* at 875 (Roberts, C.J., dissenting).

193. *See* McKusick, *supra* note 34, at 191 (noting that there has been a "growth, contemporaneously with the increase in the Court's appellate work load, of a public perception that the Court's function as the final appellate tribunal for federal questions is of overriding importance.").

V. CONCLUSION

The Supreme Court's decision in *South Carolina v. North Carolina* represents a serious misstep in the Court's original jurisdiction jurisprudence. In allowing the intervention of two nonsovereign entities in the original action, the Court overlooked the doctrine of *parens patriae* and relied upon broad cases outside the scope of equitable apportionment. The Court determined that corporations with interests adverse to a state's interests may not be adequately represented by the state, thus contravening the *parens patriae* doctrine. Further, the Court did not consider the impact of its decision on its exercise of original jurisdiction that expands the Court's awkward role as a trial court. The Court's decision has opened the doors for corporate entities to intervene, expanding an already overburdened docket and creating inconsistency in the Court's original jurisdiction jurisprudence.

CHELSEY J. HADFIELD

A TRIBUTE TO BOB LLOYD

SEMPER FIDELIS

DOUG BLAZE*

One memory in particular captures the essence of Bob Lloyd for me, both in terms of Bob's personality and his considerable contributions to the College of Law.

The 1994 Allen Novak Auction, held annually to raise money to assist destitute law students, was my first. In those days, the late April auction included all kinds of games and events on the front lawn of the law school. That year the festivities included a dunk tank prominently located on the corner of 15th Street and Cumberland Avenue. The student organizers had managed to convince Professor Lloyd to serve as the ceremonial "dunkee"—though now that I think of it, I doubt it took much convincing. At high noon, right on schedule, Bob came slowly waddling up the sidewalk along Cumberland, wearing flippers, a wet suit, a snorkel and mask, and carrying a loaded spear gun.

The students loved it. The line of students (and faculty) waiting their turn to take a shot at dunking Bob stretched up the block almost to the Carousel. At a dollar a throw, the amount of money raised that year set a new record. That image—Bob decked out in diving gear and carrying a spear gun, all done to entertain and help the students—personifies what I think is best about Bob and the law school.

For Professor Lloyd, the students always come first, just as they did when he put on his wet suit on that cold April afternoon. Because first and foremost, Bob is a teacher. He has always been completely committed to our primary mission—training future lawyers to the best of our ability. That doesn't mean that scholarship isn't important to Bob. It is. He is a very productive and influential scholar. But for Bob, scholarship is important because helps makes us better teachers and because it promotes improvement in the law we teach.

What has really set Professor Lloyd apart is his commitment to and demand for excellence. Whatever Bob does—including serving as a celebrity "dunkee"—he does to the very best of his considerable ability. I think it may be due to his Marine training, but, whatever the reason, Bob has always demanded a lot from his students and even more from himself. As a result, the students respond. Their respect for Professor Lloyd and their desire to do well in his classes are almost unmatched.

---

* Dean and Art Stolnitz and Elvin E. Overton Distinguished Professor of Law, University of Tennessee College of Law. J.D. Georgetown University Law Center, B.S. Dickinson College.

While students come first for Bob, his colleagues are not far behind. I had the good fortune to connect with Bob early in my tenure at Tennessee. Not long after I got settled in Knoxville, we met for a beer at the Old College Inn. It wasn't just a courtesy. Bob was sincerely interested in my thoughts about legal education, my new job as Director of Clinical Programs, and me as a person. He even paid for the beer. His friendship has been an important part of my satisfaction and success at Tennessee.

In fact, some of our early discussions helped lay the groundwork for Bob's work in creating our business transactions curriculum and establishing the Center for Entrepreneurial Law. As the founding director, Bob helped design Representing Enterprises, the signature course for the curriculum, to serve as the capstone experience for participating students. That course, fairly radical at the time, is now becoming more commonplace at law schools around the country.

As usual, Bob did what he did because it was good for the institution, just as traipsing across the front lawn of the law school in flippers and a mask made the Novak auction better and more successful. Whatever Bob does, he does fully and with commitment and passion. And, perhaps most important, he does it all with a keen—albeit quite dry—sense of humor. Who else would have what appears to be a motivational posture posted by his office door touting the value of "Procrastination"?

So cheers to Bob and Deanna on his retirement. But Professor Bob Lloyd will be very sorely missed. His commitment, and even his sense of humor, made the law school a much, much better place. And Novak auctions will never be the same.

THREE CHEERS FOR BOB LLOYD

BRANNON P. DENNING*

Bob Lloyd once told me that he ended up teaching rather by accident. He had decided to step off the law firm merry-go-round and become, as he put it, "as close to a ski bum as a man with a wife and family could be." One of his mentors suggested that he consider becoming a law professor instead. If Bob's success as an academic is any measure, then he undoubtedly would have set a new high—or is it a low?—for what it took to be a successful ski bum.

Certainly *my* life would have been quite different had Bob Lloyd not become a professor. He was almost single-handedly responsible for my deciding not to quit law school during the second semester of my first year. Coming off a mediocre first semester, I found myself in Bob's Contracts II class. Early in that second semester, Bob mentioned Grant Gilmore's *Ages of American Law*[1] in class, describing it as a humorous *tour d'horizon* of American law. Wandering in the stacks later, I found the book, read it, and mentioned to Bob that I had done so after one of our next classes. I remember quoting Gilmore's description of Christopher Columbus Langdell, the great Dean of the Harvard Law School, inventor of the casebook, and pioneer of the Socratic method. Langdell, Gilmore wrote, "seems to have been an essentially stupid man who, early in his life, hit on one great idea to which, thereafter, he clung with all the tenacity of genius."[2] Given my previous semester's experience with Langdell's system, I said I thought this was very funny. Bob agreed and seemed delighted that I had troubled to look up the book.

Soon thereafter, he asked me to be his research assistant. I was thrilled—not only with the money, but also at the chance to see what it was, exactly, that law professors *did* when they weren't teaching. Cite-checking Bob's articles, tracking down sources, and writing memos that provided him with footnote fodder was exciting. Reading successive drafts of his work, moreover, gave me an inkling of what it might be like to be a professor. That in turn inspired me to buckle down and work hard to overcome my first semester grades. Bob's praise and encouragement, moreover, gave me some confidence that I *could* do so, if I was willing to work. Thus Bob became one of my very important mentors.

---

* Professor of Law, Cumberland School of Law; J.D. University of Tennessee College of Law, 1995. Thanks to Ben Barton, Doug Blaze, Alli Denning, Anya East, Micki Fox, Joan Heminway, Glenn Reynolds, and Brooks Smith for comments on earlier drafts.

1. GRANT GILMORE, THE AGES OF AMERICAN LAW (1977).
2. *Id.* at 42.

During my second year, Bob constantly inquired about my job prospects and counseled me not to get discouraged. When those prospects began to improve in the fall of my third year, Bob made unsolicited calls to firms on my behalf. These carried weight. One of the partners at the firm I ultimately joined mentioned more than once how valuable Bob's recommendation was when they were deciding whether to extend me an offer.

If one is truly lucky, a mentor can become a lifelong friend. And I am most grateful for Bob's friendship over the years. He introduced me to the hilarious crime fiction of Carl Hiaasen and to John D. MacDonald's Travis McGee novels. Over lunch, we have talked about books, economics, business, and engineering—I was and am amazed that someone could know so much about so many different things. He has shared with me his experiences flying helicopters in Vietnam, practicing law at a large firm in Los Angeles, even what it is like to take up figure skating or to master statistics later in one's life. He is always interesting, lively, irreverent, and wonderful company, whether we're having a beer or he is teaching me how to wakeboard.

Finally, I would be remiss without offering an appreciation of Bob's extraordinary skill in the classroom. Problem-based pedagogy and computer-aided learning were not *au courant* in the early 1990s, but we scarcely read a case in Commercial Law. The problems in Bob's casebook forced us to read the U.C.C. closely and taught us never, ever to forget to check the definitions section of Article 9. His CALI exercise on fixtures, moreover, helped me learn that material better than a mountain of hornbooks or student aids ever could have.

Bob's passion for teaching was such that not even physical injury could keep him out of the classroom. My Contracts II class returned from Spring Break in 1993 to find Bob with his arm in a rather serious-looking sling. He had apparently dislocated his shoulder while skiing. As I recall, we were discussing damages and Bob seemed to have considerable difficulty doing three-place addition. After a couple of failed attempts, Bob sheepishly dismissed class and said we would pick it up next class. It turned out that he had taken something to . . . um . . . take the edge off the pain in his shoulder right before class. Apparently the prescription was stronger than he anticipated—strong enough, in any event, to make teaching damages as hazardous as driving or operating heavy machinery!

While Bob's retirement is a real loss for the law school, I can't help but be pleased for him. He has certainly earned the right to go, in the words of my favorite literary detective's family motto, "as [his]

whimsy takes [him]."[3] Wherever that is, I can't wait to hear about it over a pint at Sunspot.

Thanks for everything, Bob.

3. From the coat of arms of Lord Peter Wimsey, created by Dorothy Sayers. *See* DOROTHY L. SAYERS, WHOSE BODY? (1922) (introducing Lord Peter).

PILOT LLOYD

THOMAS C. GALLIGAN, JR.*

What is an eighteen-letter word or expression for Bob Lloyd? From the perspective of a law school dean getting along with[1] faculty members can be a little bit like doing crossword puzzles. When you do most crossword puzzles, the key is figuring out the theme. Sometimes the theme involves a quotation or rhyme that plays itself out over a series of clues. Sometimes it is "cuter," like familiar expressions turned around. For instance, in one puzzle I did recently the clue was something like: what professional track stars get. And the answer was: money for the runs. Of course the answer turns the expression runs for the money around. Other clues and answers did the same sort of thing with other expressions. And, those are only a couple examples of puzzle themes—crossword puzzle editors are extremely creative people.

So, how is a dean getting along with faculty like doing crossword puzzles? You have to find the theme—finding the theme or personality or character trait can be the key to a good relationship. Some colleagues are like the late, great film star, Greta Garbo; they want to be left alone. That's their theme. Others want to tell you everything—they store information and keep you informed: what are they working on; how their classes are going; what progress are their committees making; their kids need braces! Others are advice givers; they like you to sit with them in their offices—sometimes they come to your office but advice givers usually prefer their own turf—and tell you what they think of the curriculum, of their junior colleagues, of their senior colleagues, of their colleagues' scholarship, and of the overall direction of the deanship. That's their theme. Others want you to know what the mood of the faculty is; they are sometimes, dare I say it, more concerned with letting you know what their colleagues think rather than what they think. Some are just plain nervous, nervous whenever you come to see them, nervous that you did not like their last article (never!), nervous about the finances of the institution (always!), and nervous that you are about to ask them to serve on a committee. That's their theme. Of course, as a law school dean, I loved them all! And, out of necessity, I have drawn

* Thomas C. Galligan, Jr. is President and Professor of Humanities at Colby-Sawyer College in New London, New Hampshire. From 1998–2006 he was the dean of the University of Tennessee College of Law, where he also held an Elvin E. Overton Distinguished Professorship. From 1986–1998 he taught at LSU's Paul M. Hebert Law Center where he held the Dale E. Bennett Distinguished Professorship. In addition to writing about good friends, he writes and speaks about torts and admiralty.

1. The less gregarious might say "dealing with" instead of "getting along with" but that is not my view.

caricatures, none of which are based on real persons; any resemblance is merely accidental!

But, getting back to my theme, what was Bob Lloyd's theme? Bob and I met in the spring of 1998 when I was interviewing for the deanship of the University of Tennessee College of Law. As I recall, we met in his office, where he hosted a small group of faculty for a conversation about and the law school and me. That was part of how we interviewed at Tennessee, in small groups scheduled in faculty offices. It was all very welcoming and very friendly. I recall Bob had a picture of a helicopter on his wall, and I learned later that he had flown helicopters in the Viet Nam War. So maybe one theme might be "veteran," but it would only get part of it. I am sure Bob is very proud of his service but he does not proclaim it or extensively discuss it, at least not with me.

Of course Bob taught contracts and commercial law (among other things) and I had taught UCC Sales at LSU more than ten times, so we had commercial law in common. And certainly "commercial law" would be one of Bob's themes, at least from my perspective as a former commercial law teacher. But there is so much more. On the commercial/business law front, Bob was one of the inspirations for the creation of the Clayton Center for Entrepreneurial Law and served as its director. So "groundbreaking" is another theme for Bob. But my friend Dick Wirtz, who was the dean when the center was created and when Bob directed it, would be better on that score than me.

From my perspective, I would say another theme for Bob is "creative." Not only was he in on the creation of the Clayton Center for Entrepreneurial Law, but, during my time as dean, Bob designed and taught an on-line commercial law course, which he offered during two consecutive summers. It was one of the first truly on-line courses offered at an ABA law school and Bob taught it and taught it very well. So I definitely put "creative" on the theme list.

I learned early that Bob was a truly committed and successful teacher. His classes were always popular and full, and he worked constantly to make them engaging and interesting. He is devoted to teaching his students to actually practice law, and he understands how the theoretical and practical interrelate. Bob is committed to the problem method and has written and co-written leading "case" books that employ it. I would therefore also put "teacher extraordinaire" on the theme list.

Yet another aspect to Bob's teaching that is distinctive is his firm belief that law students are entering an intensely competitive practice environment in which only a small percentage of those who enter the private practice of law with law firms will ultimately become partners in those firms. As I said, Bob is determined that students understand the world of practice that they are entering—not only the law but the economic reality. He has written and spoken on the subject. So we can justifiably call Bob "realistic."

But while "realistic" about the world of practice being a tougher place than the world of law school, Bob is always willing to talk to students about

their class work or career aspirations. He is a valued advisor to those who seek him out, and his counsel is wise and directed. So he is "realistic," "kind," "available," and "wise."

Bob is not only dedicated to his teaching, his students, and his scholarship—having authored many articles in addition to the "case" books I referred to above—he is also dedicated to his family. His wife Deanna is intelligent, fun, and loaded with personality. Susan and I thoroughly enjoyed our time with her, especially at the informal UT law play reading group "productions." Bob is also extremely proud of his daughter, Christine. She was most kind to my wife and to my son, Patrick, when she was a student at Macalester College and he was a high school senior deciding where he would go to college. She met them on campus, shared her experience, and made them feel at home both with her and at Macalester. So we have to put "family man" on the Bob Lloyd theme list.

What else? Well, I was doing my daily[2] crossword the other day and the clue was: characteristic of duct tape. The answer came to me as I got a couple of the letters—it was "many uses." Indeed duct tape does have many uses and my time with Bob Lloyd truly taught me that fact. As I said, Bob flew helicopters in Viet Nam, and, a few years before I arrived at Tennessee, he took up flying again, this time planes, not helicopters. In pursuit of his flying hobby, Bob owned a partial interest in a plane, and late in August one year he asked me if I was interested in an airplane ride on Labor Day morning. I jumped at the chance, and so I met Bob very early on the appropriate Monday on the airfield by the Tennessee River on Airline Highway. I felt like I was in an old movie! The airfield is basically that—a field—grass, with planes and hangars lining the sides. As I said, I met Bob and he proceeded to introduce me to his plane, a two-seat World War II-era propeller job.

"Do you always fly so early in the morning? I asked.

"In the summer I do because of the heat and humidity," he answered.

"Because it's cooler in the morning?" I asked, clearly showing my ignorance.

"Yes, but it isn't because of comfort—it's because of air pressure and fuel burn. It's easier to fly in the morning and we'll use less fuel. We wouldn't want to run out of fuel," he said with a grin.

He did the checks you have to do before flying: he walked around the plane, looked at the wheels, and checked the fuel level. Checking the fuel level involved looking at a straightened coat hanger with a little bit of red tape on the end that protruded up and out from the fuel tank. When I asked how it worked Bob showed me how the hanger or wire was embedded in (or poked through) a cork, which floated in the fuel. Thus, when the hanger

2. Actually I probably average two per day, and on the harder ones (as the week progresses and the puzzles get more difficult), I will use the internet, even though my friend George Kuney believes that is cheating. But the point is learning, not just getting the puzzle right.

protruded out and the tape was visible at the top, the fuel was full or at least there was fuel; as the cork sank, the fuel burned. And when the wire hanger was not visible and the red tape was touching the plane's body, the tank was empty. On that Labor Day as we prepared to fly, the tank was full.

We then climbed into the cockpit—me in the front and Bob in the back. As I forced myself into my seat, I was immediately impressed (other words also come to mind) by the massive amounts of visible duct tape—hence the recent crossword puzzle reminder—which was apparent at all sorts of places where one part of the plane met another.

"Wow," I said, trying to sound nonplussed, "there's a lot of duct tape here."

"There is," said Bob, sounding as loquacious as John Wayne in "The Searchers."

"Looks good," I smiled.

"It does the job; after all it is an old plane," said Bob, returning my smile. Then he showed me how to put on the headset that we would use to communicate with one another and he told me how to use it. Next, he started up the motor, and, headsets or not, I could not hear a thing over the high-pitched scream of the engine.

We moved onto the green runway and taxied to the end. Bob turned the plane and we began to race down the grass tarmac—although I am not sure it is a tarmac if it is grass.[3] We picked up speed; I had never bounced quite so much in an airplane. It felt more like being in a small boat on rough seas. The slope of the runway seemed to be slightly uphill, so we were speeding along a rising slope; I crossed my fingers. Given the space in the cabin, my fingers were really all I could cross. Vroom, vroom, vroom, boom, boom, boom. And then we were up and off the ground. We were climbing as we crossed the Tennessee River.[4] Then, as I recall, we turned to the left.

Bob's voice came through the headset; I knew it was him because there were only the two of us and there was no AM/FM or XM radio in the plane. I strained to understand what he was saying.

"What?" I finally had to ask.

"We have to turn before we get to the Knoxville Airport," Bob replied.

"Ahhhh, makes sense," I said. The crossword solver in me even thought of the desirability of upping the "turn" to a "veer" if the occasion presented itself.

Then we headed to the West and slowly banked right and headed north, enjoying the beautiful views and flying along ridge tops. Bob told me where we were and what we were seeing. I chimed in periodically, usually to ask, "What's this?" and "What's that?" I am not sure how far north we were, but after a half-hour or so we began to make our way south again. We

---

3. Good question for someone who does crosswords.

4. I think we crossed the Tennessee River; I may be wrong. Sorry if I get the facts wrong, Bob, but crossing the river sounded good and I know we had to avoid the Knoxville Airport.

may have crossed the river and then crossed back again. I assume we avoided the Island Home takeoff and landing routes as we had avoided Knoxville Airport. At least we did not have to "veer." We were lazily heading towards UT, the law school, and Neyland Stadium.

And, then, suddenly, we veered and I mean *veered* to our left and made a beeline for Airline Highway. I wondered what I had said. Had I somehow offended my friend and host?

"What's going on?" I asked.

I assume Bob did not hear me over the screeching engine so I did not repeat my question. I was sure I had done something wrong. The plane moved towards the airfield and we descended. We descended further with nary a jolt. We were just above the green grass and then touched—or bounced—down. Actually, for such a small plane and such a pastoral runway, the landing was quite smooth. We taxied back to the place where Bob kept his plane and as I pulled off the headset and worked on releasing the straps that had held me in place, Bob literally leapt out of the cockpit.

He stepped on the wing, jumped to the ground and ran to the front part of the plane. He reached up for the spot where the wire hanger and red tape had been so apparent earlier. For the first time, I noticed that the hanger was nowhere to be seen. Bob pulled on its tip, and out came the wire and the cork, which was now near the top of the wire, not at the bottom where it had been before.

"I knew it," said Bob. "I knew we had fuel; I figured the wire had just slipped down the cork."

"You mean . . ." I stammered as I almost fell out of the cockpit and off the wing.

"Did you see it?" He asked.

"No," I said, "the cork..."

"The wire and red tape just dropped all of a sudden; it went from sticking right up there to falling down. But it just slipped, see?" He again showed me where the cork was in relation to the red tape.

"I do," I said.

He laughed; I was happy to have been so oblivious.

So, add "cautious daredevil" to the themes. I will admit my car seat felt very comfortable and secure as I sat down in it to head home for the rest of the Labor Day holiday. Bob later replaced the duct-tape plane with another one—a newer model—in which I was also fortunate to get to fly.

I am most grateful Bob got me home that first day we flew together—grateful he was cautious and got me back to the crossword business. But I am grateful to Bob for many other reasons, and they do not relate to flying. I am grateful to him for being a dedicated and wonderful teacher—in my eight years at UT, I never had to think twice about the quality of his students' educational experiences; I knew they were wonderful. Not only that, but Bob taught large sections and was willing to teach at unpopular

times on unpopular teaching days.[5] I am grateful that he always and regularly produced and published high quality, well-researched, well-written, and important scholarship.

And, I do not believe Bob ever refused me when I asked him to do something.[6] For instance, Bob agreed to and very ably chaired the search committee that led to George Kuney's appointment as Director of the Clayton Entrepreneurial Law Center. Bob willingly served on tenure and promotion committees. And, he volunteered—yes, volunteered—to chair the admissions committee. That is, he willingly served on and led heavy workload, time-consuming committees while never missing a step in the classroom and continually producing fine research.

So, what is that eighteen-letter word or expression for Bob Lloyd? "Dedicated colleague."

Thank you, Bob.

5. I think he even taught on Friday!

6. A trait he shared, by the way, with the late, great Jerry Phillips.

WHAT ABOUT BOB?[1]

JOAN MACLEOD HEMINWAY[2]

I am darn near sure that Bob Lloyd is not the kind of guy who would claim to value law review tributes. No doubt, if Bob knew that I was sitting here and writing a tribute to him, he would scoff at my heartfelt (albeit somewhat feeble) attempt to honor him as he retires. He might, in fact, tell me that I am wasting my time in writing this. I suppose I shouldn't care whether he thinks less of me for writing this; but I actually *do* care.[3]

Bob was a faithful supporter of mine as I transitioned from private practice to law teaching. He eventually served on my tenure committee and chaired my promotion committee. He read a draft of my first law review article, commenting, as I recall, that it read "like a prospectus" (which was *not* intended as a compliment, despite the 15 years I spent drafting securities offering documents). Through these and other experiences, I have come to respect (if not always agree with) his judgment. He's someone I do not want to disappoint.

Bob is almost always irreverent, sometimes even cheeky. Students and colleagues alike might be tempted to find his semi-confrontational style intimidating. (Indeed, I understand that he has asked unprepared students to leave his classroom). But he would not want to foster that view. I am confident that, instead, he would want others to interpret his impertinence as motivational, as encouragement to others to challenge assumptions and change. To the extent that he comes across as tough, it's toughness with a purpose. He believes in the College if Law as well as his students and colleagues; he wants us all to succeed.

In some of his recent work, he has begun to share this tough love with those outside the College of Law family. In *Why Every Law Student Should Be a Gunner*[4] and *Hard Law Firms and Soft Law Schools*,[5] Bob takes aim at the aspirations that law students have for themselves and the objectives that law school administrators and faculty set for their institutions and their students. But unlike others who are critical of legal education, he hasn't given up on it; rather, he's trying to improve it.

---

1. Although I am aware of the movie by the same name, *see* http://www.imdb.com/title/tt0103241/, I do not mean to make an allusion to it here. The title, however, seems apt.

2. College of Law Distinguished Professor of Law, The University of Tennessee College of Law; J.D., New York University School of Law; A.B., Brown University.

3. About now, Bob is sure to be thinking—but never would actually say—that I should "get a life." At least I was drinking a beer when I started writing this, Bob . . . .

4. Robert M. Lloyd, *Why Every Law Student Should Be a Gunner*, 40 ARIZ. ST. L.J. 1343 (2008).

5. Robert M. Lloyd, *Hard Law Firms and Soft Law Schools*, 83 N.C.L. REV. 667 (2005).

Other recent works are a testament to his desire to use his experience and intelligence to improve the substance and practice of law. In these pieces, he combines an elder-statesman-like wisdom in certain areas of the law with a broad and deep knowledge and understanding of business transactions and the people who make deals. I use some of these articles in my teaching and am citing to one in a current scholarly project. My favorites in this genre include *Discounting Lost Profits in Business Litigation: What Every Lawyer and Judge Needs to Know*,[6] *Proving Lost Profits After* Daubert*: Five Questions Every Court Should Ask Before Admitting Expert Testimony*,[7] and Pennzoil v. Texaco, *Twenty Years After: Lessons For Business Lawyers*.[8] I truly hope that Bob continues to write about important issues at the intersection of law and law practice. I find his work of this kind easy to read, credible, and useful.

I know Bob would want me to mention briefly (in passing) a few of his sartorial choices that I have admired over the past ten-plus years (since no one else likely will raise this point, especially in a serious law review tribute). After all, we don't just talk about *law* in the faculty wing! Bob has his own unique fashion sense; I might describe his personal style as post-military preppy. He is one of few adult males that I have seen in or around Knoxville (another being, as I recall, our colleague Ben Barton) wearing Nantucket Reds,™[9] plain-front or pleated cotton canvas trousers in an unusual, faded/muted red. The days he wears his Reds to the office bring me back to the Northeast mentally and emotionally. And Bob already knows this, but he really should wear pink oxford cloth shirts more often. They are great with his coloring and (as we all have agreed on the third floor) do not compromise his manliness in the slightest.

In case you can't tell, I will miss Bob's counsel, intensity, support, encouragement, intelligence, and (yes) personal style. But, as with all good people who have worked long and hard hours for the folks and institutions they love, Bob must be set free to enjoy and impact others. He leaves us with a better place by his presence and his many, many contributions. And for all the demurrals that may issue from Bob in response to the collegial reflections published in this volume, I somehow believe that he will appreciate in his heart the homages we all have written to and of him. As well he should.

---

6. Robert M. Lloyd, *Discounting Lost Profits in Business Litigation: What Every Lawyer and Judge Needs to Know*, 9 TRANSACTIONS: TENN. J. BUS. L. 9 (2007)

7. Robert M. Lloyd, *Proving Lost Profits After* Daubert*: Five Questions Every Court Should Ask Before Admitting Expert Testimony*, 41 U. RICH. L. REV. 379 (2007).

8. Robert M. Lloyd, Pennzoil v. Texaco, *Twenty Years After: Lessons For Business Lawyers*, 6 TRANSACTIONS: TENN. J. BUS. L. 321 (2005).

9. *See* http://nantucketreds.com/mens/pants/reds.html.

TENNESSEE LAW REVIEW
Member of the National Conference of Law Reviews

Volume 78 Spring 2011 Number 3

CONTENTS

FOREWORD: DIVINE OPERATING SYSTEM?
Glenn Harlan Reynolds 651

ARTICLE V CONSTITUTIONAL CONVENTIONS: A PRIMER
Editors of the Tennessee Law Review 663

ARTICLES

PROPOSING CONSTITUTIONAL AMENDMENTS BY CONVENTION: RULES GOVERNING THE PROCESS
Robert G. Natelson 693

SOVEREIGNTY, REBALANCED: THE TEA PARTY & CONSTITUTIONAL AMENDMENTS
Elizabeth Price Foley 751

REOPENING THE CONSTITUTIONAL ROAD TO REFORM: TOWARD A SAFEGUARDED ARTICLE V CONVENTION
Michael Stern 765

CONVENTIONAL WISDOM: ACKNOWLEDGING UNCERTAINTY IN THE UNKNOWN
Mary Margaret Penrose 789

A BRIEF REPLY TO PROFESSOR PENROSE
Michael Stern 807

PROPOSED AMENDMENTS

THE CASE FOR THE REPEAL AMENDMENT
Randy E. Barnett 813

AMENDING ARTICLE V TO MAKE THE CONSTITUTIONAL AMENDMENT PROCESS ITSELF LESS ONEROUS
Timothy Lynch 823

THE TRUTH-IN-LEGISLATION AMENDMENT: AN IDEA WHOSE TIME HAS COME
Brannon P. Denning & Brooks R. Smith 831

AN EQUAL RIGHTS AMENDMENT TO MAKE WOMEN HUMAN
Ann Bartow 839

WHY WE NEED TERM LIMITS FOR CONGRESS: FOUR IN THE SENATE, TEN IN THE HOUSE
Richard A. Epstein 849

THE EXECUTION SHOULD BE TELEVISED: AN AMENDMENT MAKING EXECUTIONS PUBLIC
David Lat & Zachary Shemtob 859

AFTERWORD: FULL OF SOUND AND FURY BUT SIGNIFYING RELATIVELY LITTLE?
Sanford Levinson 867

FOREWORD: DIVINE OPERATING SYSTEM?

GLENN HARLAN REYNOLDS*

Discussion of constitutional amendments—and especially a full-blown constitutional convention—seems inevitably to bring the Beatles' "You say you'll change the Constitution?" line to mind.[1] But in pondering this Symposium, I was put in mind of another musical reference: "Divine Operating System."[2] This is because the Constitution is regarded by many on both the left and the right as being of at least somewhat divine provenance, and because the role that it serves in our system of government is more like that of an operating system than of software. But as any computer user can attest, operating systems do not notably partake of the divine.

One of the most over-used and under-appreciated statements in constitutional law is Chief Justice John Marshall's admonition to remember that "we must never forget, that it is *a constitution* we are expounding."[3] Philip Kurland observed more recently that when modern judges cite this statement, "you can be sure that the Court will be throwing the constitutional text, its history, and its structure to the winds in reaching its conclusion."[4] There is some truth to this observation, but it is more a reflection of modern judges than on Chief Justice Marshall's statement. Marshall's statement instead captured the difference between a constitution—which lays down structure and general rules but cannot "partake of the prolixity of a legal code"[5]—and, well, the prolixity of a legal code. Ignore this distinction and you get something like the California constitution, which can certainly be so described, but which few regard as a model of elegant or effective drafting.

I. THE OPERATING SYSTEM MODEL

A computer operating system, like a constitution, establishes a structure and a set of basic rules. That structure has certain consequences for what comes later—compare Windows with OS X with Ubuntu—but in general, the purpose of the operating system is to establish the framework, while fleshing out that structure with function is left to the application software that is layered on top. What the operating system permits software to do is

---

* Beauchamp Brogan Distinguished Professor of Law, The University of Tennessee.

1. THE BEATLES, *Revolution*, *on* THE BEATLES (Apple 1968).

2. SUPREME BEINGS OF LEISURE, DIVINE OPERATING SYSTEM (Palm Pictures 2002).

3. McCulloch v. Maryland, 17 U.S. 316, 407 (1819).

4. Philip B. Kurland, *Curia Regis: Some Comments on the Divine Right of Kings and Courts "To Say What the Law Is,"* 23 ARIZ. L. REV. 581, 591 (1981).

5. *McCulloch*, 17 U.S. at 407.

important, but so are the things that software is prohibited from doing. Flaws in those prohibitions can lead to unnecessary conflicts, memory leaks, and catastrophic misallocation of resources. Those flaws are (sometimes) fixed in later updates.

In principle, of course, the operating system can do anything the software layered on top of it does, and to a large degree the reverse is true as well. You could design an operating system that contained a full-featured word processor and spreadsheet, and you could produce software that compensates for operating-system deficiencies. But it is much better to keep the two separate, as the operating system does a better job of keeping watch on the applications programs if the two are not combined.

Though I am a fan of constitutional metaphors in general,[6] the idea of an operating system metaphor for the Constitution is not entirely my own. It is explored by Neal Stephenson in his history of operating systems and their consequences, *In The Beginning . . . Was The Command Line*, where he warns that abandoning the constraints of an established constitutional operating system is much more dangerous than trying out the latest iteration of Windows on your laptop. Stephenson writes that in the twentieth century:

> [I]ntellectualism failed, and everyone knows it. In places like Russia and Germany, the common people agreed to loosen their grip on traditional folkways, mores, and religion, and let the intellectuals run with the ball, and they screwed everything up and turned the century into an abattoir. Those wordy intellectuals used to be merely tedious; now they seem kind of dangerous as well. We Americans are the only ones who didn't get creamed at some point during all of this. We are free and prosperous because we have inherited political and value systems fabricated by a particular set of eighteenth-century intellectuals who happened to get it right. But we have lost touch with those intellectuals.[7]

Stephenson's cautionary note is one that advocates of wholesale constitutional revision—or perhaps even modest constitutional change—should bear in mind. Though sweeping and/or clever changes to the Constitution can make for enjoyable seminar discussion, the machinery of government is subject to catastrophic failure (failure which, based on human history, seems to be more-or-less inevitable on any significant time scale). Those who would tinker with that machinery need to be very aware of the potential damage that can result from unwise alterations, as the term

---

6. *See, e.g.*, Glenn Harlan Reynolds, *Chaos and the Court*, 91 COLUM. L. REV. 110 (1991) [hereinafter Reynolds, *Chaos and the Court*] (chaos theory as model for Supreme Court decisionmaking); Glenn Harlan Reynolds, *Is Democracy Like Sex?* 48 VAND. L. REV. 1635 (1995) [hereinafter Reynolds, *Is Democracy Like Sex?*] (comparing the role of democracy in restraining special-interest networks with the role of sexual reproduction in boosting resistance to parasitism).

7. NEAL STEPHENSON, IN THE BEGINNING . . . WAS THE COMMAND LINE 53 (1999).

"catastrophic" is no metaphor here. In this sense, I think that constitutional amendments are perilous, and a constitutional convention is analogous to the "hyperspace" button in the old *Asteroids* videogame—worth pressing only *in extremis*, since it was almost as likely to kill you as to save you.

On the other hand, the constitutional operating system is not divine—it was the product of human beings, who were subject to human blindnesses and failures, and like any human creation it is subject to being rendered obsolescent by future developments. It was not, in Michael Kammen's memorable phrase, intended to be a "machine that would go of itself."[8] The Framers inserted the Article V amendment process out of an expectation that their work would require adjustments over time. The adoption of the Constitution itself was accompanied by an enormous ten-amendment project known as the Bill of Rights. Amendments Eleven and Twelve followed quite soon after, as imperfections in the original scheme were discovered. The Reconstruction Amendments followed after considerable difficulty revealed other substantial flaws. Furthermore, of course, the Constitution has been effectively amended by judicial interpretations on numerous occasions as well. If the Republic can face the risks of amendment via judicial action with equanimity, it can surely face the risks inherent in amendment via the procedures of Article V.

So if constitutional amendments, large and small, are in a sense part of the system, how can we determine what sorts of amendments are best?

II. AMENDMENTS GOOD AND BAD

Predicting the result of even modest changes in a complex system is often difficult, and sometimes impossible.[9] But my first observation is that a constitutional amendment should be directed at something structural, something basic to the functioning of our government, and not at some particular policy issue or legislative question. If one is possessed of a supermajority but fears losing it in the future, there may be a powerful temptation to "lock in" policy preferences by promoting them to the level of the Constitution, where future change to undo them will be much more difficult. But such changes are not really the objective of the amendment process.

And, indeed, a survey of the constitutional amendments adopted to date tends to support the wisdom of this approach. Most are structural reallocations of power: between the states and the federal government (and between black people and white people) in the Reconstruction Amendments, between the federal government and individuals in the Sixteenth Amendment, between the people and states in the Seventeenth Amendment, between men and women in the Nineteenth Amendment,

---

8. MICHAEL KAMMEN, A MACHINE THAT WOULD GO OF ITSELF: THE CONSTITUTION IN AMERICAN CULTURE (1986).

9. *See generally* Reynolds, *Chaos and the Court*, *supra* note 6.

between the executive and the electorate in the Twenty-second Amendment, between Congress and the electorate in the Twenty-eighth. Others fix structural problems: reducing dangerous "lame duck" periods in the Twentieth Amendment and fixing bugs in presidential succession in the Twenty-fifth Amendment.

In fact, the only major departure from this principle is to be found in the most unsuccessful constitutional amendment of all time, the repealed 18th Amendment, which imposed an alcohol prohibition. Rather than being primarily about structure, it was primarily about policy. It was also a spectacular failure. This experience should, perhaps, encourage caution regarding similar efforts. Note that this involves both form and substance: An amendment to ban abortion would seem to resemble Prohibition, while an amendment to place abortion decision-making exclusively in the states would not; likewise a law to ban gay marriage versus placing the decision in the hands of state legislatures.

At any rate, I offer some (brief) thoughts to follow on a number of proposals for constitutional change and reform beyond those that are spelled out at greater length—and, no doubt, greater erudition—by the participants in this Symposium. While many if not all of them probably belong on whatever constitutes the constitutional equivalent of the cutting-room floor, they will perhaps serve as a spur to further thought. In all cases, I will try to observe the constitution-as-operating-system distinction between constitutional subjects and legislative software. Readers may judge how well I succeed.

In addition, I proceed on the basis that, with the federal government at record size and the federal debt at record levels, constitutional amendments intended to bring runaway government expansion under control are most likely to interest the electorate over the coming years. I may be wrong, but I believe that social-issue amendments, on topics like abortion or gay marriage, are unlikely to generate the sort of interest they might have in less parlous financial circumstances. Certainly it seems likely that, should a constitutional convention be called in the near future, it will be to deal with these sorts of financial and structural problems, not the social-issue bugbears of yore.

III. CH-CH-CH-CHANGES

A. Taking Some Amendments Seriously

In the past I have proposed altering the Ninth Amendment to read: "The enumeration, in this Constitution, of certain rights, shall not be construed so as to deny or disparage other rights retained by the people. *And we really mean it!*"

That, I thought, would make clear that the Constitution's distribution of power consists of discrete government powers in a sea of individual rights, and not the other way around. The only change I would make today would

be to add the same postscript to the Tenth Amendment as well. With limited government unpopular among politicians—because, one suspects, it is insufficiently productive of graft—and with federal courts largely unwilling since the New Deal to police the "internal" limits of government power, some additional external constraints might well be worth it. A somewhat more serious—though less amusing—version of this might provide, instead of *And we really mean it!* that no law made in violation of such principles could be upheld, and that any U.S. citizen should have standing to seek judicial remedies of violations. But perhaps reliance on the federal courts to protect liberties is a mistake. And so, the next proposal.

B. Setting Ambition Against Ambition in Cause of Smaller Government: A House of Repeal

Our structure of government, in which laws must pass bicamerally and then be presented to the President, and, after passage, face judicial review from the courts, would seem to produce powerful checks against the overweening growth of government power. This, however, has turned out not to be the case—or, at least, such checks have not proven powerful enough.

Thus, in keeping with James Madison's principle of letting ambition counteract ambition, I suggest a third house of Congress whose sole function is to repeal laws. Members should serve staggered terms, and their only power—and hence, the only thing they could place before constituents in seeking reelection—would be to repeal laws enacted by Congress.[10]

There might be concerns about infinite recursion—in which Congress passes bills, has them repealed, and then re-passes them to face a further repeal—though in practice I wonder how likely that might be (or whether it would be so terrible, if it were to happen). However, there could be limitations, such as a requirement that the House of Repeal act by a two-thirds vote of each house, or a provision that a House of Repeal could only repeal a given law once per session, or a veto-style override provision allowing the House and Senate to re-pass a "repealed" law by a two-thirds margin in each house. (Or, perhaps more sensibly, forbidding them from re-passing a repealed law until an election has intervened). Perhaps even all of these might apply, or with different standards depending on the age or subject matter of the legislation in question, depending on how much repeal one believes is likely to be necessary.

Though the details, as with all constitutional provisions, matter a lot, the key virtue of a House of Repeal goes beyond the details: The point of its existence would be to give *someone* in the federal government an incentive

10. This idea is not original with me. As far as I know, it was first aired in Robert Heinlein's novel *The Moon Is a Harsh Mistress*. *See* Dmitri Feofanov, *Luna Law: The Libertarian Vision in Heinlein's* The Moon Is A Harsh Mistress, 63 TENN. L. REV. 71, 126–27 (1995).

to give us *less* law rather than more. Right now, only the federal judiciary is free from incentives to create additional regulation (though not necessarily free from incentives to create additional legal complexity),[11] but federal judges get no reward for striking laws down. There is no institutional incentive to do so. Yet it seems that things are much more likely to get done in our system if some institution benefits from the doing.

In a sense, the House of Repeal proposal would simply institutionalize Randy Barnett's Repeal Amendment proposal[12] under one roof. The difference is that while state legislatures might sometimes favor repealing particular legal provisions, they too lack an institutional incentive to identify poorly functioning federal laws and eliminate them. In our current system, no one has such an incentive, really. This proposal would remedy that.

C. No Representation Without Taxation

Since the adoption of the Sixteenth Amendment and the institution of a progressive income tax, we have seen an enormous expansion of federal spending and an increasing inability of the political system to keep such spending from growing. There are numerous reasons for this phenomenon, but one characteristic of a progressive income tax system is that people who make less money have less of a stake in controlling spending.

Indeed, according to a 2009 report from the Tax Foundation, the income tax burden of the top one percent of taxpayers exceeds that of the bottom ninety-five percent.[13] The top fifty percent of earners, meanwhile, paid over ninety-seven percent of income taxes (and the top twenty-five percent paid over eighty-six percent), while the bottom fifty percent paid less than three percent, and many not only paid no taxes but actually received money back due to various refundable-credit schemes for low-income workers.[14] One need not be an opponent of progressive taxation in general to recognize that such a narrow concentration of tax burdens poses risks in a largely majoritarian political system. Those who pay no taxes—as a large number of Americans do not[15]—or whose taxes are negligible as a

---

11. *See, e.g.*, Benjamin H. Barton, *Do Judges Systematically Favor the Interests of the Legal Profession?*, 59 ALA. L. REV. 453 (2008).

12. *See generally* Randy E. Barnett, *The Case for the Repeal Amendment*, 78 TENN. L. REV. 813 (2011); Randy E. Barnett & William J. Howell, *The Case for a 'Repeal Amendment'*, WALL ST. J., Sept. 16, 2010, at A23.

13. Scott A. Hodge, *Tax Burden of Top 1% Now Exceeds That of Bottom 95%*, TAX POLICY BLOG (July 29, 2009), http://www.taxfoundation.org/blog/show/24944.html.

14. Gerald Prante & Mark Robyn, *Fiscal Fact: Summary of Latest Federal Individual Income Tax Data*, TAX FOUND., 3 (Oct. 6, 2010), http://www.taxfoundation.org/files/ff249.pdf.

15. Stephen Ohlemacher, *Nearly Half of U.S. Households Escape Fed Income Tax*, YAHOO! FINANCE (Apr. 7, 2010, 5:38 PM), http://finance.yahoo.com/news/Nearly-half-of-

fraction of their income, are likely to be rather sanguine about the burdens of increased spending and increased taxation, since those burdens will fall on others.[16]

The American Revolution was based, in part, on cries of "no taxation without representation," but the problem here is something a bit different. Those who pay substantial taxes get to vote, but the progressivity of the system ensures that they will be in the minority. And it seems no great stretch to foresee a future in which less than half the electorate will pay any federal income tax at all.

A system in which voting for increased spending and taxation has no visible cost to the majority of voters is a system that is likely to suffer from much-higher-than-optimal levels of spending and taxation. Perhaps some sort of corrective is in order.

The flip side of "no taxation without representation"—"no representation without taxation"—is questionable. Such an approach, limiting the franchise to those who pay income taxes, would likely violate the Twenty-fourth Amendment, which provides:

> The right of citizens of the United States to vote in any primary or other election for President or Vice President, for electors for President or Vice President, or for Senator or Representative in Congress, shall not be denied or abridged by the United States or any State by reason of failure to pay any poll tax or other tax.[17]

And at any rate, shrinking the scope of the franchise is not likely to promote government stability or accountability.[18] On the other hand, there is no reason why the Sixteenth Amendment could not be amended to limit the progressivity of the federal income tax. Under an ideal system, everyone, regardless of income, would pay at least some income tax (enough to notice—say in the neighborhood of five percent of gross personal income), and the amount paid would fluctuate up or down in tandem with federal spending. More spending should hurt, at least a little. But how do we keep Congress from making up the difference through borrowing, as it has done evermore frequently lately? That brings us to the next proposal.

US-households-apf-1105567323.html?x=0&.v=1.

16. The percentage of households paying no income tax climbed rapidly in recent years. Tad DeHaven notes that "[a]s the price of something drops, the demand increases. For a growing share of Americans, government services are effectively 'free,' so they are demanding even more and policymakers are giving it to them." Tad DeHaven, *The Something-for-nothing Quandary*, CATO@LIBERTY (Sept. 15, 2010, 3:00 PM), http://www.cato-at-liberty.org/the-something-for-nothing-quandary.

17. U.S. CONST. amend. XXIV, § 1.

18. In his future history series, science fiction writer Jerry Pournelle has envisioned a future in which America is divided between dole-receiving "Citizens" and more elite "Taxpayers." It is not an especially appealing future. *See, e.g.*, JERRY POURNELLE, HIGH JUSTICE (1979).

D. A Question of Balance

Nearly every state in the Union already possesses some sort of state-constitutional provision requiring a balanced budget, though the rigor of these provisions varies.[19] There is no corresponding provision in the federal Constitution, whose Article I, Section 8 merely authorizes Congress "[t]o borrow Money on the credit of the United States."[20]

Many have considered that to be a defect and have proposed various flavors of balanced budget amendments designed to curb federal borrowing. Here is the text of one version, introduced in the 111th Congress:

> Section 1. Total outlays for any fiscal year shall not exceed total receipts for that fiscal year, unless three-fifths of the whole number of each House of Congress shall provide by law for a specific excess of outlays over receipts by a rollcall vote.
> Section 2. The limit on the debt of the United States held by the public shall not be increased, unless three-fifths of the whole number of each House shall provide by law for such an increase by a rollcall vote.
> Section 3. Prior to each fiscal year, the President shall transmit to the Congress a proposed budget for the United States Government for that fiscal year in which total outlays do not exceed total receipts.
> Section 4. No bill to increase revenue shall become law unless approved by a majority of the whole number of each House by a rollcall vote.
> Section 5. The Congress may waive the provisions of this article for any fiscal year in which a declaration of war is in effect. The provisions of this article may be waived for any fiscal year in which the United States is engaged in military conflict which causes an imminent and serious military threat to national security and is so declared by a joint resolution, adopted by a majority of the whole number of each House, which becomes law.
> Section 6. The Congress shall enforce and implement this article by appropriate legislation, which may rely on estimates of outlays and receipts.
> Section 7. Total receipts shall include all receipts of the United States Government except those derived from borrowing. Total outlays shall include all outlays of the United States Government except for those for repayment of debt principal.
> Section 8. This article shall take effect beginning with the later of the second fiscal year beginning after its ratification or the first fiscal year beginning after December 31, 2014.[21]

---

19. *See generally* NAT'L CONFERENCE OF STATE LEGISLATURES, *NCSL Fiscal Brief: State Balanced Budget Provisions* (Oct. 2010), http://www.ncsl.org/documents/fiscal/StateBalancedBudgetProvisions2010.pdf (providing constitutional and statutory citations for state balanced budget requirements).

20. U.S. CONST. art. 1, § 8.

21. H.R.J. Res. 1, 111th Cong., (2009).

This version found little traction and the 111th Congress was not notable for its progress toward balanced budgets. But note that this language is milder than many state balanced-budget provisions, in that it provides for unbalanced budgets by a relatively small (three-fifths rather than two-thirds) supermajority. It also does not limit the assumption of unfunded future liabilities, something that has proven to be a serious problem for many states.[22] It is likely to be a problem for the federal government as well, and any serious balanced budget proposal should take such liabilities into account.

Another approach would grant the President a line-item veto whenever the budget is out of balance.[23] This is even weaker tea. Of course, at the opposite end of the spectrum, it is possible to imagine extending Article I, Section 10's prohibition against the states making "any Thing but gold and silver Coin a Tender in Payment of Debts,"[24] to the federal government, but that seems, well, rather strong tea.

Still, the demonstrated tendency of politicians to keep on spending other people's money well past the point at which it has run out suggests that some sort of check might be worthwhile. A survey of state balanced budget proposals—which, in some cases, as in Oregon, limit the size of surpluses as well as deficits[25]—might be worthwhile. Because states cannot print their own money, the problem of overspending has come up more often, and the solutions they have arrived at are likely to have considerable relevance now that the federal government is approaching its limits.

E. Turnover

Many Americans interested in political reform are unhappy with the low rate of turnover in Congress. With its long-term civil service employment practices, the bureaucracy, of course, hardly turns over at all. But Congress is little better. Even in national elections where the shift seems drastic, like the midterm elections of 2010, the vast majority of incumbents are still re-elected. This has led to a number of proposals for limiting the terms of representatives and senators, both at the federal level[26]—where, unsurprisingly, they have found limited congressional

---

22. *See, e.g.*, Megan McArdle, *Dire States*, THE ATLANTIC, Jan./Feb. 2011, at 36 (describing budgetary difficulties of many states resulting from unfunded pension liabilities, etc.).

23. *See* Anthony W. Hawks, *The Balanced Budget Veto: A New Mechanism to Limit Federal Spending*, POL'Y ANALYSIS, Sept. 4, 2003, at 1.

24. U.S. CONST. art. 1, § 10.

25. *See* OREGON DEP'T OF REVENUE, OREGON CORPORATE EXCISE AND INCOME TAX: CHARACTERISTICS OF CORPORATE TAXPAYERS APP. D (2007), *available at* www.oregon.gov /DOR/STATS/docs/102-405-FY07/102-405-07.pdf.

26. For a current term-limits proposal, see generally Richard A. Epstein, *Why We Need Term Limits for Congress: Four in the Senate, Ten in the House*, 78 TENN. L. REV. 849

support—and at the state level, where the Supreme Court, in *U.S. Term Limits v. Thornton*,[27] found such beyond state power.

Were we to hold a constitutional convention, of course, congressional defensiveness would no longer be a roadblock, and, I suppose, it is because of such potential roadblocks to change that the convention approach was included as an alternative to amendment by congressional action. That said, I remain skeptical that mechanical term limits are the way to go. In my essay, *Is Democracy Like Sex?*, I discussed the role of electoral turnover in limiting special-interest corruption and suggested that a term limits approach would be much less effective:

> The imposition of mandatory turnover on elective offices certainly tends to change things around, but it is not at all certain that it would accomplish as much as the reshuffling brought about by democratic electoral politics. The value of "shuffling," after all, is that it is more or less random. The turnover created by term limits would not be random at all. In addition, the term-limit remedy acts whether it is needed or not. Turnover accomplished by electoral processes, on the other hand, may be in part "random," but it may also stem—as I think it has in the last couple of elections—from a widespread sense on the part of voters that special interest parasitism has gotten out of hand. Even if we do not feel that we can count on voters to engage in the kind of day-to-day effort required to make plebiscitary democracy work—something hard to expect in an age when we cannot get people to show up for jury duty—perhaps we *can* count on them to know when things have gotten too cozy, and to act appropriately. Certainly the end of the Cold War has produced just such a sense, and just such action, in quite a few democracies besides this one.[28]

I continue to feel this way, though I confess that, in the intervening years, I have become more concerned that the role of gerrymandering may be undercutting the prospects for democratic turnover to a greater degree than I appreciated. To the extent that this is true, a constitutional amendment aimed at limiting gerrymandering might do more good than term limits. Personally, I would rather see an official leave office because the voters kicked the individual out, than simply because of the calendar.

V. CONCLUSION

The other participants in this symposium have far more to contribute than I have managed here, but I hope that these observations will serve to spur discussion. And, should the nation choose to try a second constitutional convention, I hope that my words of caution will encourage

---

(2011).

27. 514 U.S. 779 (1995) (finding unconstitutional state-imposed term limits on federal representatives).

28. Reynolds, *Is Democracy Like Sex?*, *supra* note 6, at 1650 (citations omitted).

some degree of additional care and consideration in any participants who happen to read them. When our Constitution was drafted, the participants in that convention knew that they were undertaking something audacious, unprecedented, and risky, but, under the circumstances, they considered the risk worth running in the hopes of producing a stable, prosperous, and free society that would last many decades. That they succeeded beyond their wildest dreams should not be taken as a reason for any successors to be less careful.

ARTICLE V CONSTITUTIONAL CONVENTIONS: A PRIMER*

I. INTRODUCTION 664
II. THE HISTORICAL FATE OF ARTICLE V CONVENTIONS 665
A. *Constitutional Convention Close Calls* 666
1. Reapportionment 666
2. Balanced Budget 667
B. *The Historical Absence of Article V Conventions* 668
III. ARTICLE V AND THE PRESIDENCY 669
IV. CONGRESS'S ROLE IN A CONVENTION 670
A. *Congressional Power to Call (or Refuse to Call) a Convention* 670
1. The Scope of a Convention's Deliberations 671
a. *The Unlimited Convention View* 671
b. *The Limited Convention View* 673
i. Single Amendments 674
ii. Specific Subject Areas 674
2. Congressional Review of Applications for a Limited Convention 674
3. Timeliness of Applications 675
4. Effect of Rescinded Applications 677
B. *Post-Call and Post-Convention Powers* 678
V. THE ROLE OF THE SUPREME COURT IN A CONVENTION 678
A. *The Political Question Doctrine and an Article V Convention* 680
B. *Justiciability and Article V* 680
C. *Judicial Activism* 683
VI. THE ORGANIZATION OF AN ARTICLE V CONVENTION 685
A. *The Calling of a Convention* 685
1. Forum Selection 686
2. State Representation 686
3. Selection of Delegates 688
B. *The Operation of a Convention* 689
VII. CONCLUSION 691

* Sara R. Ellis, Yusuf Z. Malik, Heather Graves Parker, Benjamin C. Signer, & Al'Reco L. Yancy, candidates for Doctor of Jurisprudence, University of Tennessee College of Law, May 2011.

The authors would like to thank Audrey M. Calkins and Ryan Connor for their assistance in compiling this primer.

The Congress, whenever two thirds of both houses shall deem it necessary, shall propose amendments to this Constitution, or, on the application of the legislatures of two thirds of the several states, shall call a convention for proposing amendments, which, in either case, shall be valid to all intents and purposes, as part of this Constitution, when ratified by the legislatures of three fourths of the several states, or by conventions in three fourths thereof, as the one or the other mode of ratification may be proposed by the Congress; provided that no amendment which may be made prior to the year one thousand eight hundred and eight shall in any manner affect the first and fourth clauses in the ninth section of the first article; and that no state, without its consent, shall be deprived of its equal suffrage in the Senate.[1]

I. INTRODUCTION

Article V of the United States Constitution provides two procedures for amending the Constitution.[2] First, Congress may propose amendments when they are approved by at least a two-thirds majority in both the House of Representatives and the Senate.[3] Thereafter, if three-fourths of the states ratify the proposed amendment, it becomes part of the Constitution.[4] To date, this method has been employed every time the Constitution has been amended.[5] Second, Article V prescribes an alternative method of amendment in which Congress calls a constitutional convention upon applications by two-thirds—currently thirty-four—of the states.[6] After the delegates to the convention propose an amendment, it becomes part of the Constitution upon approval by three-fourths of the states, just as in the congressional-proposal amendment procedure.[7]

Uncertainty suffuses the notion of holding an Article V convention in our time; the only other constitutional convention took place over two centuries ago, and the text of Article V offers limited guidance.[8] Despite decades of scholarly attempts to define a convention's contours and clarify convention procedures, scholarly unanimity has not emerged on key points.

---

1. U.S. CONST. art. V.

2. *See id.*; *see also* James Kenneth Rogers, Note, *The Other Way to Amend the Constitution: The Article V Constitutional Convention Amendment Process*, 30 HARV. J.L. & PUB. POL'Y 1005, 1005 (2007).

3. *See* U.S. CONST. art. V; Rogers, *supra* note 2, at 1005.

4. *See* U.S. CONST. art. V; Rogers, *supra* note 2, at 1005.

5. *See* Rogers, *supra* note 2, at 1005 ("The second method . . . has never been used.").

6. *See* U.S. CONST. art. V; Rogers, *supra* note 2, at 1005.

7. *See* U.S. CONST. art. V.

8. *See* Gerald Gunther, *The Convention Method of Amending the United States Constitution*, 14 GA. L. REV. 1, 2 (1979) (noting that the last constitutional convention took place in Philadelphia in 1787); Rogers, *supra* note 2, at 1005.

Indeed, there is even disagreement over the level of disagreement that is warranted.[9]

Given such uncertainty, this primer will provide a general overview of many of the issues surrounding an Article V convention rather than definite answers to remaining questions about the convention process. Indeed, certainty simply cannot be the aim of any Article V discussion given the paucity of available guidance on the subject and the extant scholarly disagreement.[10] Moreover, a comprehensive analysis of all of the procedural issues confronting a convention is beyond the scope of this introductory piece. We seek merely to make the reader generally familiar with major issues implicated by Article V's convention method.

First, we look at some historical instances when constitutional conventions have nearly been called and offer some possible explanations for why no convention has been called. Second, even though the President of the United States has no formal role in amending the Constitution, we discuss the possible sideline impact the White House could have on a convention. Third, we explore Congress's role in a modern convention: its responsibilities and potential limitations on its involvement. Fourth, we turn to the United States Supreme Court's role with respect to the convention alternative, including the Court's potential to referee a convention, the influence the political question doctrine may have on judicial review of convention questions, and the potential for charges of "judicial activism" should the Court enter the political thicket. Finally, we discuss the potential organization of a modern constitutional convention, including the selection of its forum, choice of delegates, and the procedural operation of the convention.

II. THE HISTORICAL FATE OF ARTICLE V CONVENTIONS

Despite the submission of approximately 750 applications for an Article V convention, including applications by all fifty states, no constitutional

---

9. *Compare* Neal S. Manne, Note, *Good Intentions, New Inventions, and Article V Constitutional Conventions*, 58 TEX. L. REV. 131, 135 (1981) ("For the constitutional law scholar, the consideration of the convention alternative is a foray into conjecture and speculation. The method has never been used, and the text, history, and policy considerations relating to the convention are all less than unambiguous."), *and* Gunther, *supra* note 8, at 20 (recognizing "the large number of unresolved questions posed by the constitutional convention route"), *with* Thomas Brennan, *The Last Prerogative*, 6 HARV. J.L. & PUB. POL'Y 61, 69 ("All the procedural questions can be answered by reference to the words of Article V, the nature of the convention and the historical precedent of the constitutional convention in Philadelphia in 1787. There is no need for any new rules to guide the convention. The convention can organize itself, and perform its historic function just as its predecessors at both the state and national levels have done.").

10. *See* Gunther, *supra* note 8, at 11 ("[N]o one can make absolutely confident assertions about how the convention method was intended to operate.").

convention has ever been called.[11] This section will explore two instances when the country came very close to holding a convention: the reapportionment debate of the mid-twentieth century and the balanced budget crisis of the 1970s and 1980s. Then, it will speculate as to why a modern convention has never been called.

A. Constitutional Convention Close Calls

Since the 1960s, the thirty-four-state threshold required to call an Article V convention has nearly been met on two occasions.[12] The first such instance concerned reapportionment issues in the 1950s, while the second involved calls for a balanced budget in the latter half of the twentieth century.[13]

1. Reapportionment

Issues surrounding the apportionment of votes and voting districts created quite a furor in the states in the mid-twentieth century.[14] In *Wesberry v. Sanders*, the Supreme Court struck down Georgia's apportionment of its congressional districts as discriminatory.[15] Similarly, in *Reynolds v. Sims*, the Court ordered reapportionment of Alabama's legislative districts.[16] In response to these decisions, state legislatures filed

---

11. *See* FRIENDS OF THE ARTICLE V CONVENTION, *Article V Amendment Applications Tables*, http://foa5c.org/file.php/1/Articles/AmendmentsTables.htm (last visited May 5, 2011) [hereinafter *Applications Tables*]; *see also* Dwight W. Connely, Note, *Amending the Constitution: Is This Any Way to Call for a Constitutional Convention?*, 22 ARIZ. L. REV. 1011, 1011–12 (1980).

12. *See* Rogers, *supra* note 2, at 1009.

13. *See id.*

14. *See id.*

15. *See* Wesberry v. Sanders, 376 U.S. 1, 7 (1964). Citizens of Georgia's Fifth Congressional District claimed that the size of their district was disproportionate to Georgia's other voting districts, thereby giving their votes unequal weight compared to the votes of citizens of other districts. *See id.* at 2–3. In fact, the Fifth Congressional District was over twice the size of the average district in Georgia and three times larger than the smallest district. *Id.* In holding that Georgia's apportionment "grossly discriminate[d] against voters in the Fifth Congressional District," the Court stated that "[i]f the [f]ederal Constitution intends that when qualified voters elect members of Congress each vote be given as much weight as any other vote, then [the] statute cannot stand." *Id.* at 7.

16. *See* Reynolds v. Sims, 377 U.S. 533, 586 (1964). A group of Alabama citizens challenged the apportionment of Alabama's legislature under the Equal Protection Clause of the Fourteenth Amendment. *Id.* at 537. The districts were created using the population of the 1900 federal census, and the population had grown exponentially by the time of the lawsuit. *Id.* at 540. In ordering reapportionment of Alabama's legislative districts, the Court held that the apportionment of the legislature must be based on population and that "an individual's right to vote for state legislators is unconstitutionally impaired when its weight is in a

applications to address the issue of reapportionment through the convention mechanism.[17] Within five years of *Wesberry* and *Reynolds*, thirty-three states submitted applications, just one state shy of the requisite thirty-four.[18] However, as the movement lost momentum, several states withdrew their applications.[19] There are two explanations for the failure to reach thirty-four applications: first, the realization by reapportionment opponents that the Supreme Court's decisions would not adversely affect traditional rural interests,[20] and second, the mounting fear among some that a convention would not be limited to the reapportionment debate but would instead ignore the original scope of its call and overhaul the Constitution.[21] In short, although the states failed to reach the requisite number of applications, they came quite close.

2. Balanced Budget

A more recent attempt at an Article V convention came when thirty-two states submitted applications calling for an amendment requiring a balanced budget in response to the increasing federal deficit in the 1970s and 1980s.[22] Although no convention was called, the rapid submission of applications put enormous pressure on Congress to confront the deficit issue.

Confronted by the threat of a convention, Congress passed the Gramm-Rudman-Hollings Balanced Budget and Emergency Deficit Control Act of 1985, which required a balanced budget by 1993.[23] The Act purported to accomplish exactly what a constitutional amendment would have—a balanced budget—without the expenditure of resources and the inherent unpredictability of a convention.[24] Although no convention was called,

---

substantial fashion diluted when compared with votes of citizens living in other parts of the State." *Id.* at 568. Thus, *Reynolds* "laid down the basic proposition that state legislators should be elected from districts of equal population (or proportional population in the case of multimember districts)." Robert G. Dixon, Jr., *Article V: The Comatose Article of Our Living Constitution?*, 66 MICH. L. REV. 931, 933–34 (1967).

17. *See* Rogers, *supra* note 2, at 1009.

18. *See* RUSSELL L. CAPLAN, CONSTITUTIONAL BRINKSMANSHIP 76 (1988); Rogers, *supra* note 2, at 1009.

19. *See* Rogers, *supra* note 2, at 1009. However, the states' rescissions may have had no effect had Congress decided to call a convention.

20. *See* CAPLAN, *supra* note 18, at 76; Rogers, *supra* note 2, at 1009.

21. *See* CAPLAN, *supra* note 18, at 76.

22. *See* Anthony James Guida, Jr., Case Note, *States' Role in Article V Conventions:* AFL-CIO v. Eu, 54 U. CIN. L. REV. 317, 317 (1985). Concerns about the extensive national debt, partially caused by an unbalanced budget, are also at the forefront of today's political debate.

23. *See* 2 U.S.C. §§ 900–922 (2006). For a detailed discussion of the passage and revision of the Gramm-Rudman-Hollings Act, see CAPLAN, *supra* note 18, at 84–89.

24. Of course, an amendment would have had a more lasting effect than the Act. If a

states successfully used the threat of the Article V alternative as a vehicle for political reform.[25]

B. The Historical Absence of Article V Conventions

There are two likely reasons that an Article V convention has not been called despite nearly 225 years of availability. First, Congress has at least some power to interpret the Article's requirements, allowing it some control over when its constitutional duty to call a convention has been triggered.[26] Second, Congress has preempted potential convention calls through action designed to address the concerns that given rise to a successful call.

Although over 750 Article V applications have been filed by the states since our nation's last constitutional convention,[27] Congress has yet to call a convention; thus, it is reasonable to assume that Article V includes a subject-specific requirement that the states have yet to satisfy.[28] For example, thirty-six states proposed 167 amendments from 1963 to 1969, and thirty-eight states proposed fifty-seven amendments between 1965 and 1971.[29] However, during both periods, states' applications covered a multitude of subjects rather than a single specific topic,[30] and Congress's duty to call a convention may not have been triggered. Yet, while thirty-eight states have submitted 123 balanced budget amendment applications

---

convention had been called, perhaps the United States would not be in the same budgetary position it is today.

25. *See* John T. Noonan, Jr., *The Convention Method of Constitutional Amendment—Its Meaning, Usefulness, and Wisdom*, 10 PAC. L.J. 641, 645 (1979) ("A reform that strikes at the power of Congress may only be adopted if effective pressure is generated by the states. The way of generating effective pressure is the way provided by the founding fathers—application by the legislatures of the states for a Convention."); Michael B. Rappaport, *Reforming Article V: The Problems Created by the National Convention Method and How to Fix Them*, 96 VA. L. REV. 1509, 1535 (2010) ("[T]he threat of a convention may cause Congress to pass an amendment that it otherwise would have refused to enact.").

26. *See infra* Part IV.A.

27. *See Applications Tables*, *supra* note 11.

28. Caplan posits that "applications must evidently specify particular amendments, and a convention need be called only if the requisite number of applications agree in text or subject matter with regard to at least one amendment." CAPLAN, *supra* note 18, at 99. However, Caplan also notes that Congress "has considered, but never passed, constitutional convention implementation acts whose main purpose is to limit conventions to the subject matter contained in the applications." *Id.* at 94. Regardless, it is reasonable to assume that at least thirty-four specific applications must share the same subject area. If not, Congress has neglected its duty to call a convention for many years. *See* Michael Stokes Paulsen, *A General Theory of Article V: The Constitutional Lessons of the Twenty-Seventh Amendment*, 103 YALE L.J. 677, 736 (1993).

29. *See Applications Tables*, *supra* note 11, at 24–26. This excludes any rescissions made by the states. *Id.*

30. *See id.* at 24–26.

since 1955,[31] satisfying the two-thirds requirement and the subject matter requirement, they may have failed to meet an implied "reasonable time" requirement.[32]

It is possible that fear of the uncertainties accompanying a convention (or of the diminution of its power that could result from one) has caused Congress simply to exercise its power of application review so as to prevent a convention from occurring. As was partially the case in the reapportionment debate, the threat of an uncontrollable convention that "could not be limited to a single subject" may have stalled the convention drive.[33] Accordingly, calls for a convention seem either to resolve themselves on their own, as in the calls to address reapportionment in the mid-twentieth century, or conjure so much political anxiety that the desired result is achieved through more conventional means, as in the balanced budget debate of the 1970s and 1980s.[34]

III. ARTICLE V AND THE PRESIDENCY

Article V makes no mention of any presidential role in either method of amendment it provides,[35] and the Supreme Court has ruled expressly that the President has no official role in the constitutional amendment process. In *Hollingsworth v. Virginia*, Attorney General Lee mentioned in oral argument that "the case of amendments is evidently a substantive act, unconnected with the ordinary business of legislation, and not within the policy, or terms, of investing the President with a qualified negative on the acts and resolutions of Congress."[36] Justice Chase agreed with this reasoning, stating that "[t]he negative of the President applies only to the ordinary cases of legislation: He has nothing to do with the proposition, or

---

31. *See id.* at 26.

32. *See* CAPLAN, *supra* note 18, at 110–14 (discussing the "timeliness" requirement in detail). *But see* Paulsen, *supra* note 28, at 735 n.197 (rejecting an implicit time requirement).

33. Rogers, *supra* note 2, at 1009–10. Theoretically, a general convention could rewrite our current Constitution, greatly destabilizing the state of the law. See Rappaport, *supra* note 25, at 1528–31, for a detailed discussion of the potential problems associated with a "runaway convention." *But see* Brennan, *supra* note 9, at 68 ("There is no danger of a runaway convention."); Noonan, *supra* note 25, at 642–43 ("There is absolutely nothing . . . that suggests that the states can only call a general Convention where the whole Constitution will be on the table to be bargained over."); Paulsen, *supra* note 28, at 742 ("Those who dread a 'runaway' convention thus misapprehend the very nature of a constitutional convention, which is inherently illimitable in what it may propose.").

34. *See* Rogers, *supra* note 2, at 1009–10.

35. *See* U.S. CONST. art. 5.

36. Hollingsworth v. Virginia, 3 U.S. (3 Dall.) 378, 381 (1798). In *Hollingsworth*, opponents of the Eleventh Amendment argued that it was invalid, because it was not presented to the President before going to the states. *Id.* at 379. The Court rejected this argument and held that the "amendment [had been] constitutionally adopted." *Id.* at 382.

adoption, of amendments to the Constitution."[37] Consequently, *Hollingsworth* declares that the President has no veto power over a proposed constitutional amendment and is neither required nor permitted to take action for an amendment to become effective under Article V.[38] And while this case was decided in the context of the congressional-proposal method of amendment, there is no reason to suppose that this decision would not extend with equal force in the convention-method context.[39]

Of course, the President's lack of an explicit role in the amendment process does not mean that he or she would have no influence on it; Presidents exert influence over members of their party and would no doubt do so in the context of an Article V convention. But, as the foregoing indicates, there is apparently no place for direct presidential intervention.

IV. CONGRESS'S ROLE IN A CONVENTION

Article V provides: "The Congress . . . on the application of the Legislatures of two thirds of the several States, shall call a Convention for proposing Amendments . . . ." But precisely how this is to play out is the source of great controversy. The devil, as always, is in the details. Some scholars contend that the procedures are straightforward and that congressional involvement in a convention would be minimal.[40] Others contend that the relationship between Congress and an Article V convention is murky at best and that no clear answers to it are readily visible.[41] In this section, we will attempt to address some of the potential issues surrounding Congress's involvement in the convention process.

A. Congressional Power to Call (or Refuse to Call) a Convention

Article V provides that Congress *shall* call a Convention for proposing amendments on the application of two-thirds of the states.[42] On a plain reading of this language, Congress appears mandated to call a convention upon such application.[43] Yet, although hundreds of Article V applications

37. *Hollingsworth*, 3 U.S. (3 Dall.) at 381 n.*.

38. *Id.* Two scholars, Homer Clark and Laurence Tribe, agree. *See* Homer Clark, *The Supreme Court and the Amending Process*, 39 VA. L. REV. 621, 623 (1953); Laurence Tribe, *Issues Raised by Requesting Congress to Call a Constitutional Convention to Propose a Balanced Budget Amendment*, 10 PAC. L.J. 627, 634 (1979).

39. *See* CAPLAN, *supra* note 18, at 168 ("[A]mendments proposed by . . . a convention are not subject to the [presidential] veto . . . .").

40. *See, e.g.*, Brennan, *supra* note 9, at 66–69.

41. *See, e.g.*, Laurence Tribe, *supra* note 38, at 634–40.

42. *See* U.S. CONST., art. V (emphasis added).

43. As Noonan has pointed out, Alexander Hamilton regarded Congress's duty to call a convention upon the application of two-thirds of the states "'peremptory.'" Noonan, *supra* note 25, at 642 (quoting THE FEDERALIST No. 85 (Alexander Hamilton)); *see also* CAPLAN, *supra* note 18, at 116 ("A few writers agree that even if the requisite applications were filed,

have been filed by the states since our nation's last constitutional convention,[44] Congress has yet to call a convention. Therefore, either Congress has unconstitutionally refused to honor Article V's mandate or the applications themselves have somehow failed to trigger this obligation. Caplan has argued that "Article V implies that Congress is the agent entrusted to receive, inspect, and decide on the validity of applications."[45] In the absence of widespread accusations of such congressional dereliction, it is reasonable to assume that Congress simply has not found that the aggregation has triggered its constitutional duty. But precisely when and how an application individually "counts" toward the two-thirds requirement is unclear, as is when applications collectively would mandate a convention call. This depends on a number of factors, including the extent to which Congress can regulate the scope of a convention, the timeliness of the applications, and the effect of state rescission of applications. Additionally, Congress can, and has, acted legislatively to preempt the concerns that motivate states to call for conventions in the first place.

1. The Scope of a Convention's Deliberations

Of central importance is whether Congress (or the states) may constitutionally limit the scope of a convention's deliberations or whether a convention, once called, attains the plenary authority to propose amendments on subjects not explicitly contemplated in states' applications.

a. The Unlimited Convention View

One school of thought holds it constitutionally impermissible to limit the scope of a convention's deliberations. Michael Rappaport has argued recently that "the greater part of the commentary appears to argue for the unlimited convention view."[46] There are several lines of argument in its favor. First, as Rappaport notes, Article V's text contains no plain

---

Congress may refuse a call, but most scholarly opinion finds an inescapable obligation."); Douglas G. Voegler, *Amending the Constitution by the Article V Convention Method*, 55 N.D. L. REV. 355, 368 (1979) ("The evidence that Article V places a mandatory duty upon Congress to call a convention, when properly petitioned, is overwhelming.").

44. *See Applications Tables*, *supra* note 11; *see also* Connely, *supra* note 11, at 1011–12.

45. CAPLAN, *supra* note 18, at 94; *see also* Voegler, *supra* note 43, at 369 ("Congress . . . has the power to ascertain whether the prerequisites to [its duty to call a convention] have been met."); Connely, *supra* note 11, at 1017–20.

46. Rappaport, *supra* note 25, at 1518; *see also* Paulsen, *supra* note 28, at 738 ("[T]here can be no such thing as a 'limited' constitutional convention. A constitutional convention, once called, is a free agency."); Voegler, *supra* note 43, at 384 ("The authorities overwhelmingly believe that the states [have no power to limit the scope of an Article V convention].").

limitations on the convention's operations.[47] Second, it has been argued that a convention becomes essentially a sovereign, acquires for the duration of its existence plenary authority over its deliberations, and is subject neither to congressional nor state control.[48] Third, Walter Dellinger has argued that the Framers intended the convention method of amendment to provide an alternative to the congressional method of amendment that would also prevent the states from colluding to amend the Constitution self-interestedly at the expense of the nation.[49] On this theory, the convention scheme was designed to interpose the "substantive involvement of a national forum" between the states' proposal and ratification of amendments.[50] Needless to say, an unlimited convention would have remarkable power to fundamentally change the American government.[51]

If the requisite number of states apply explicitly for an unlimited convention, it seems likely that Congress will be constitutionally mandated to call an unlimited convention.[52] Less clear is what would happen if the requisite number of states applied did not *explicitly* apply for an unlimited convention. According to Rappaport, whether the constitutional duty is triggered in this case

> depend[s] on the nature of the applications. If it is clear that the state applications request only a limited convention, then Congress would not even be authorized to call a convention, because there would be no applications for the only constitutional type of convention—an unlimited convention. By contrast, if the states preferred a limited convention but were understood to be applying for an unlimited convention if a limited convention were not deemed legal, then Congress would be required to call an unlimited convention and any proposals it would make would be legal.[53]

---

47. *See* Rappaport, *supra* note 25, at 1519.

48. For a discussion of this view, see Connely, *supra* note 11, at 1021.

49. Walter E. Dellinger, *The Recurring Question of the "Limited" Constitutional Convention*, 88 YALE L.J. 1623, 1626 (1979).

50. *Id.* at 1630.

51. Of course, even an unlimited convention would be limited by political reality: anything proposed would have to secure ratification by three-fourths of the states.

52. *See* CAPLAN, *supra* note 18, at x ("The evidence indicates that . . . plenary conventions . . . may be applied for . . . .").

53. Rappaport, *supra* note 25, at 1518 (footnotes omitted); *see also* Dellinger, *supra* note 49, at 1626 ("If those thirty-four states *recommend* in their applications that the convention consider only a particular subject, Congress must still call a convention and leave to the convention the ultimate determination [of the agenda and the amendments to be proposed]. . . . [But if] a state's application is based on the erroneous assumption that Congress is empowered to impose subject-matter limits on the convention, such an application must be considered invalid.") (emphasis added).

In this latter case, the drafting of the convention application would be crucial, and the states could potentially secure an unlimited convention by including language indicating their desire for one should their application for a limited convention be deemed unconstitutional.

b. The Limited Convention View

In contrast with those interpreting Article V to authorize only unlimited conventions, proponents of the limited convention view claim that the states can prospectively limit the scope of a convention's operations by specifying narrower subject matter in their applications to Congress, which will be obligated to call a convention for some limited purpose. Despite his finding that the unlimited convention view holds in the greater amount of scholarship, Rappaport maintains that there are also strong reasons supporting the limited convention view. First, he notes the language of the article can plausibly be read to cover both limited and unlimited conventions.[54]

Second, he argues there are at least three strong reasons to interpret the structure and purpose of Article V to indicate the Framers intended to authorize limited as well as unlimited conventions. First, he wonders "why would the constitutional enactors allow the states to decide not to hold any convention—and thereby to determine that none of the current problems warrant a convention—but not allow them the lesser power of determining that only certain problems warrant a convention?"[55] In other words, the lesser power follows *a fortiori* from the greater. Second, there is simply little reason to think that the Framers would have intended deprive state legislatures of the power to call for limited, as well as unlimited, conventions.[56] Finally, permitting states to apply for a convention to consider only a single amendment, the narrowest type of limited convention possible, would not remove the convention's power of proposal because the convention itself would still "decide whether the specific amendment should be proposed . . . . [T]he states could propose nothing [by themselves]."[57]

Assuming the limited convention view is constitutionally permissible, it remains unclear what level of similarity between the applications would be necessary to trigger Congress's duty to call a convention.

54. *See* Rappaport, *supra* note 25, at 1519–20.
55. *Id.* at 1521.
56. *See id.*
57. *Id.*

i. Single Amendments

One possibility is that Congress would consider only applications containing draft language of a proposed amendment.[58] Caplan, for instance, has argued that to the extent that limited conventions are constitutionally permissible, "[the convention] is bound by Article V to propose only those amendments described in the triggering applications."[59] If this were the case, the convention itself would amount to little more than a simple up-or-down vote on the amendment advanced by the state legislatures. Such a strict limitation would mean that states desirous of a convention would have to coordinate their efforts carefully to ensure that their applications contained identical or substantially similar language.[60]

ii. Specific Subject Areas

Another possibility is that a convention could be limited to the particular subject area set out in the states' applications.[61] For example, the states could submit applications calling for a convention to deal with some particular issue or set of issues. This more liberal approach would give convention delegates more flexibility to fully debate the substance and language of proposed amendments, as well as alter the language of the states' proposals as necessary. Such an approach would be advantageous in giving convention delegates the freedom necessary to address the states' concerns while at the same time preventing them from embarking on an amendatory process contrary to the states' desires.[62]

2. Congressional Review of Applications for a Limited Convention

Whether the single-amendment or subject-matter limitation holds, in the event of applications for some type of limited convention, Congress would have more latitude (and difficulty) in deciding whether to call a convention in the first instance and potentially in exercising authority over its activities once it had been called.[63]

---

58. *See* Dellinger, *supra* note 49, at 1631.

59. CAPLAN, *supra* note 18, at x.

60. Such a narrow view of the convention's authority is anathema those like Dellinger, who argue that such control by the state legislatures and Congress would rob the convention of its essential character. *See* Dellinger, *supra* note 49, at 1630–31.

61. *See* Rappaport, *supra* note 25, at 1518.

62. Even Dellinger concedes that "[i]t is possible that a set of state applications could establish subject matter limitations sufficiently broad to provide latitude for compromise and consensus-building at the convention and sufficiently uniform to enable Congress to define and enforce those limits without unduly intruding into the convention's work." *See* Dellinger, *supra* note 49, at 1635–36.

63. For a discussion of issues surrounding the convention's organization and

Rappaport notes that Congress would face several potential difficulties in deciding whether the applications for a limited convention triggered its duty.[64] This is especially the case if "the states [sought] conventions on similar but not identical subjects[.]"[65] In reading such applications, Congress would confront "a host of . . . difficult questions."[66] Of course, if Congress were hostile to a convention, its power to evaluate applications could also provide it with a pretext not to call one.

3. Timeliness of Applications

In addition to determining whether the applications are constitutional in scope, Congress might also consider their timeliness—how long must Congress consider an application in its tally once it has been submitted? Article V does not include language establishing temporal deadlines for filing the requisite number of state applications or for the ratification process after an amendment has been proposed.[67] Fortunately, two Supreme Court cases indicate that such time constraints likely exist with respect to ratification and potentially exist with respect to applications.

First, in *Dillon v. Gloss*, the Court reasoned that Congress has the authority to provide for such a deadline in the ratification context.[68] The Court stated that nothing in Article V "suggest[ed] that an amendment once proposed is to be open to ratification for all time, or that ratification in some of the States may be separated from that in others by many years and yet be effective."[69] Even though Article V does not contain an express time limit for the proposal and subsequent ratification of an amendment, the Court interpreted the text as suggesting that "[the two events] are not to be widely separated in time."[70] Notably, the Court chose to go no further than this; instead, it simply ruled that the seven-year time limit associated with the adoption of the Eighteenth Amendment was reasonable.[71]

Although this reasonable time requirement arose within the context of a congressionally proposed constitutional amendment, it makes sense to apply the same requirement to the amendment convention method.

---

congressional involvement in it, see *infra* Part VI.

64. *See* Rappaport, *supra* note 25, at 1522–23.

65. *Id.* at 1522.

66. *Id.* at 1523.

67. *See* U.S. CONST. art. V.

68. Dillon v. Gloss, 256 U.S. 368, 374–75 (1921). In *Dillon*, the Court discussed the Eighteenth Amendment, which contained a provision requiring ratification within seven years of submission. *Id.* at 371–72.

69. *Id.* at 374. Thus, applying *Dillon* to the application process, Congress may impose time constraints "during which state applications [can be] considered together." Voegler, *supra* note 43, at 371.

70. *Dillon*, 256 U.S. at 374–75.

71. *Id.* at 375–76. The Twenty-First Amendment contained a similar provision. U.S. CONST. amend. XXI, § 3.

However, the *Dillon* Court's reluctance to define the limits of a "reasonable time" and the subsequent ambiguity that accompanies this decision[72] could be a source of future litigation in either amendment process. While the reasonable time requirement remains somewhat amorphous, *Dillon* does indicate that, within in the context of a convention, Congress has the power to establish the controlling time limit for ratification and that a seven-year limit may be deemed reasonable.[73]

In *Coleman v. Miller*, the Court invoked the political question doctrine and declined to address the reasonableness of the thirteen years the Kansas state legislature took to ratify the Child Labor Amendment.[74] In doing so, the Court provided two holdings relevant to our discussion. The Court first held that the efficacy of a proposed constitutional amendment's ratification by state legislatures is a political question and thus nonjusticiable, "with the ultimate authority in the Congress in the exercise of its control over the promulgation of the adoption of the amendment."[75] In its opinion, the Court revisited *Dillon* and reasoned that although Congress has the power to fix a reasonable time for the ratification of an amendment, its failure to exercise such power does not require the Court to do so.[76] In declining to review whether the Kansas state legislature's delay was reasonable, the Court noted that "the question, what is a reasonable time, lies within the congressional province"[77] and that this question was not at issue in *Dillon*.[78] Despite its reluctance to address the issue, the Court did point to several factors that Congress could consider to determine whether the time between an amendment's proposal and its ratification is reasonable, including "an appraisal of a great variety of relevant [political, social, and economic] conditions."[79]

Together, *Dillon* and *Coleman* stand for the proposition that Congress may, but is not required to, fix a reasonable time within which the states must ratify amendments proposed at an Article V convention. Furthermore,

---

72. *Dillon*, 256 U.S. at 376.

73. *Id.* The ratification must occur within a reasonable time, and Congress may fix that length of time. *Id.* at 375–76.

74. Coleman v. Miller, 307 U.S. 433, 450 (1939). In *Coleman*, the Court examined the validity of the Kansas legislature's ratification of the Child Labor Amendment ("CLA"). *Id.* at 435–36. The Kansas legislature rejected the CLA eighteen months after Congress proposed it. *Id.* at 435. However, thirteen years later, Kansas ratified the proposed CLA. *Id.* at 436.

75. *Id.* at 450. "In determining whether a question falls within [the political question and nonjusticiable] category, the appropriateness under our system of government of attributing finality to the action of the political departments and also the lack of satisfactory criteria for a judicial determination are dominant considerations." *Id.* at 454–55.

76. *Id.* at 452–53; *see* Note, *Proposing Amendments to the United States Constitution by Convention*, 70 HARV. L. REV. 1067, 1069 (1957) [hereinafter *Proposing Amendments*].

77. *Coleman*, 307 U.S. at 454.

78. *Id.* at 453.

79. *Id.*

although a reviewing court could invoke the political question doctrine and decline to review timeliness issues arising from Congress's receipt of applications for a convention, they must likely be submitted within a reasonable time of one another as defined by Congress or interpreted by the courts.

4. Effect of Rescinded Applications

Also possibly affecting Congress's responsibility to count an application is whether a state's rescission of that application nullifies it. The validity of rescinded applications cannot be divined from the text of Article V, which does not expressly grant or deny states the right to rescind an application once it has been submitted to Congress.[80]

Fortunately, the events behind the ratification of the Fourteenth Amendment may provide an answer as to whether a state's rescission of a convention application is valid.[81] In 1868, Ohio and New Jersey initially ratified the Fourteenth Amendment but later withdrew their consent.[82] After their withdrawal, Congress requested the Secretary of State to compile a list detailing which state legislatures had ratified the Amendment.[83] In his report, the Secretary noted the withdrawals of Ohio and New Jersey, yet doubted their validity.[84] Accordingly, "[o]n the following day the Congress adopted a concurrent resolution which . . . declared the Fourteenth Amendment to be a part of the Constitution and that it should be duly promulgated as such by the Secretary of State."[85] The list of ratifying states included both Ohio and New Jersey.[86] Following this logic, a state's decision to rescind its vote in favor of ratification of a congressionally proposed amendment is likely ineffectual for purposes of Article V.[87]

However, it remains unsettled what effect a state's rescission of a convention application would have. These rescissions occurred in the context of ratification, not application. Moreover, there is earlier evidence contradicting the congressional action taken in 1868.[88] Nevertheless, Caplan argues that while Congress has never addressed this question explicitly, "the power to retract a ratification was accepted by the founders, suggesting the same holds true for applications" and "the weight of the

---

80. *See* U.S. CONST. art. V. In fact, the Constitution contains no mention of the states' ability to rescind applications for a convention. *Id.*

81. *See Coleman*, 307 U.S. at 448–50.

82. *See id.* at 448.

83. *See id.*

84. *See id.* at 448–49.

85. *Id.* at 449 (citation omitted).

86. *Id.*

87. *Id.*

88. CAPLAN, *supra* note 18, at 109–10.

evidence . . . supports the power of states to rescind an application as well as a ratification."[89]

B. Post-Call and Post-Convention Powers

Congress's power and influence in the Article V convention process do not cease with the commencement of a convention. First, to the extent the limited convention view is correct, the convention cannot permissibly exceed the scope of the call from the states that brought it into being.[90] Accordingly, Congress could have authority to enforce this prerogative by reviewing the proposals produced by convention delegates to determine whether they are consistent with the original call.[91] If Congress determines that a convention's work is inconsistent with that call, it might be able to void the convention itself or any state attempts to ratify the proposed amendment.[92] Caplan, for example, has suggested that any amendments proposed by the convention going beyond its authorization "[could] be withheld by Congress from ratification and . . . [could], for the most part, be challenged in the federal courts."[93] Moreover, Congress likely has final authority over any controversies that arise from the ratification process following the convention.[94] Finally, the extent to which Congress can control the mechanics of a convention remains unclear whether the convention is limited or unlimited.[95]

V. THE ROLE OF THE SUPREME COURT IN A CONVENTION

The United States Supreme Court has had occasion to interpret several aspects of Article V in the last century.[96] These interpretations, while

---

89. *Id.* at 108, 110.

90. *See* Noonan, *supra* note 25, at 642.

91. *See id.* at 644.

92. *But see* Dellinger, *supra* note 49, at 1631 ("[The role of defining the convention's subject matter] should be left to the convention itself in order to avoid undue congressional influence over the convention mode of amendment."); Paulsen, *supra* note 28, at 738 ("A constitutional convention . . . is inherently illimitable in what it may propose.").

93. CAPLAN, *supra* note 18, at x. However, Caplan believes that there is an implicit statute of limitations on such challenges, and that "irregularly adopted amendment[s] . . . not contested over a period of years, probably decades, [could] attain a secure place in the Constitution by virtue of public acquiescence." *Id.*

94. *See* EDWIN MEESE III, THE HERITAGE GUIDE TO THE CONSTITUTION 286 (2006); *see also* U.S. CONST. art. V. ("[T]he one or the other Mode of Ratification may be proposed by the Congress . . . ."); Rappaport, *supra* note 25, at 1526.

95. *See* Rappaport, *supra* note 25, at 1523–26. For a discussion of some possibilities about a convention's mechanics, see *infra* Part VI.

96. *See generally* Coleman v. Miller, 307 U.S. 433 (1939) (recognizing Article V as the only provision for amendment of the Constitution); Smiley v. Holm, 285 U.S. 355 (1932) (recognizing the multiple functions of legislators in interpreting the meaning of

limited in scope, provide some guidance for how the Court (and lower courts) might address various issues surrounding an Article V convention.[97] The term "Legislature," Congress's requirement to call a convention, the necessity of agreement and involvement throughout the nation, the prohibition of the President's involvement in a convention, and the timeliness of an amendment's proposal and ratification are all issues the Court has addressed.[98] Although these interpretations stem from issues concerning the more traditional method of constitutional amendment, it is reasonable to believe they may be relevant in the context of an Article V convention, especially given the structure of Article V's text.[99]

As observed above, Article V does not provide much guidance as to how such a convention should operate.[100] While the Constitution gives Congress the power to call the convention but provides no explicit role for any other branch of the federal government,[101] it stands to reason that the U.S. Supreme Court could play a large role in resolving the conflicts that would almost certainly emerge should a convention be called.[102]

Some scholars argue that the Supreme Court should have no place in the convention amendment process because no constitutionally mandated role for the courts exists.[103] Nevertheless, as it seems nearly inevitable that conflict will crop up at some point in the convention process, it may be, as Tribe has argued, that "the Supreme Court would almost certainly be asked to serve as referee."[104] Faced with such a role, the Court would have to determine if the conflict presents a justiciable question.[105] If the Court decides to enter the fray, it may face familiar criticism for its perceived judicial activism.

"Legislature"); Hawke v. Smith, 253 U.S. 221 (1920) (defining "Legislature" under Article V).

97. Indeed, only three facts are "known or knowable" in regard to the convention method of amendment: (1) Congress has a duty to call the convention; (2) amendments proposed and ratified will become part of the Constitution; and (3) "the President has no role to play in the amendment process." Tribe, *supra* note 38, at 634.

98. *See supra* notes 96–97.

99. *See supra* note 96; *see also* U.S. CONST. art. V.

100. *See* U.S. CONST. art. V.

101. Clark, *supra* note 38, at 622.

102. *See* Tribe, *supra* note 38, at 634–37. Because the Supreme Court's role is not addressed by Article V, its role is not "known or knowable," leaving open the possibility that the Court could play a role in the amendment process. *Id.* at 634.

103. *See* Thomas E. Baker, *Towards a "More Perfect Union": Some Thoughts on Amending the Constitution*, 10 WIDENER J. PUB. L. 1, 6 (2000); Thomas Millet, *The Supreme Court, Political Questions, and Article V—A Case for Judicial Restraint*, 23 SANTA CLARA L. REV. 745, 748 (1983); *Proposing Amendments*, *supra* note 76, at 1068.

104. Tribe, *supra* note 38, at 636 (discussing the Supreme Court's role as referee "[i]n the event of a dispute between Congress and the Convention over the congressional role in permitting the convention to proceed . . . .").

105. *See id.*

A. The Political Question Doctrine and an Article V Convention

The Supreme Court's self-imposed policy of judicial restraint, partially embodied by the political question doctrine, could keep it from playing a role in a controversy involving an Article V convention.[106] Still, the Court has not yet decided the justiciability of specific questions surrounding a convention, and there are certainly other situations in which the Court initially appeared to refrain from entering a politically charged debate only to change course and enter the fray.[107]

This judicial policy of non-review may seem contrary to courts' lack of hesitation to intervene in other areas. But the political question doctrine recognizes that there are some constitutional issues inappropriate for judicial resolution.[108] Courts determine whether a case falls under the political question doctrine by examining the issue to see if it falls into one of several categories, "including foreign relations, questions involving dates of duration of hostilities, the formal validity of legislative enactments, the status of the Indian tribes, and questions about whether a republican form of government exists in the states"[109]; if so, the issue qualifies as a political question and is likely nonjusticiable. But the political question doctrine is not susceptible to easy categorization, and courts routinely determine whether it applies on a case-by-case basis.[110]

B. Justiciability and Article V

In *Coleman v. Miller*,[111] the Supreme Court stopped just short of foreclosing judicial review of the amendment process. In a concurrence,

---

106. *See* Thomas E. Baker, *Exercising the Amendment Power to Disapprove of Supreme Court Decisions: A Proposal for a "Republican Veto,"* 22 HASTINGS CONST. L.Q. 325, 339 (1995).

107. *See generally* Bush v. Gore, 531 U.S. 98, 111 (2000). The Court, in an unsigned opinion, said:

> None are more conscious of the vital limits on judicial authority than are the Members of this Court, and none stand more in admiration of the Constitution's design to leave the selection of the President to the people, through their legislatures, and to the political sphere. When contending parties invoke the process of the courts, however, it becomes our unsought responsibility to resolve the federal and constitutional issues the judicial system has been forced to confront.

Id.

108. JESSE H. CHOPER ET AL., LEADING CASES IN CONSTITUTIONAL LAW 17 (2009); *see also* Baker, *supra* note 106, at 339.

109. Baker v. Carr, 369 U.S. 186, 210–11 (1962); CHOPER ET AL., *supra* note 108, at 17.

110. *Baker*, 369 U.S. at 210–11.

111. 307 U.S. 433 (1939).

Justice Black, writing for four justices, stated that "[t]he process [of amendment] itself is 'political' in its entirety, from submission until an amendment becomes part of the Constitution, and is not subject to judicial guidance, control or interference at any point."[112] Although Justice Black's opinion was merely a concurrence, it serves as an example of how the Court might approach a controversy involving Article V. Further insights into the Court's amendment jurisprudence can be gleaned from several pre-*Coleman* cases, in which the court decided several amendment-related issues on the merits,[113] but it remains unclear whether the Court will deem amendment-related issues political and thus nonjusticiable.

The Court's decision in *Baker v. Carr* attempts to clarify the question of justiciability.[114] The *Baker* Court set out the following standard in an attempt to clarify the definition of a "nonjusticiable political question":

> Prominent on the surface of any case held to involve a political question is found a textually demonstrable constitutional commitment of the issue to a coordinate political department; or a lack of judicially discoverable and manageable standards for resolving it; or the impossibility of deciding without an initial policy determination of a kind clearly for nonjudicial discretion; or the impossibility of a court's undertaking independent resolution without expressing lack of the respect due coordinate branches of government; or an unusual need for unquestioning adherence to a political decision already made; or the potentiality of embarrassment from multifarious pronouncements by various departments on one question.[115]

Assuming the Court applies this definition to a modern convention, it could help create a role for itself in the amendment proceedings because, as the Court stated in *Baker*, "[u]nless one of these formulations is *inextricable* from the case at bar, there should be no dismissal for nonjusticiability on the ground of a political question's presence."[116]

On the other hand, *Coleman* may also indicate that the Court would decline to interfere in a convention's operations. Some scholars put more weight onto Justice Black's *Coleman* concurrence, contending that it and other Supreme Court decisions strongly suggest that "Article V places the

---

112. *Id.* at 459 (Black, J., concurring).

113. *See, e.g.*, Leser v. Garnett, 258 U.S. 130, 136 (1922); Dillon v. Gloss, 256 U.S. 368, 374 (1921); *see also* Note, *Proposed Legislation on the Convention Method of Amending the United States Constitution*, 85 HARV. L. REV. 1612, 1636 (1972) [hereinafter *Convention Method*] ("[I]n *Dillon v. Gloss* the Court ruled that seven years was a "reasonable" time limit for Congress to place on ratification of an amendment. And in *Leser v. Garnett*, the Court held that state legislatures, in ratifying proposed amendments, could not be bound by procedural requirements imposed by their state constitutions.").

114. *Baker*, 369 U.S. at 217.

115. *Id.*

116. *Id.* (emphasis added). For an in-depth discussion of each of the *Baker* factors and their application to Article V issues, see *Convention Method*, *supra* note 113, at 1637–41.

primary responsibility for amending the Constitution within the province of the legislative branch, so the courts should play no role whatsoever in the process of considering amendments."[117] Further, as Millet has noted, "at no time has the Court ever exercised its power of judicial review to invalidate an amendment or any step taken by Congress in the amendment process."[118] Moreover, the Court has consistently held that there are "no implicit limits on the content of amendments that may be proposed and ratified, thus evidencing the seeming tautology that a provision properly added to the Constitution cannot be unconstitutional."[119]

The Court's deference to Congress, particularly evident in *Coleman*, may have been partially due to the political climate at the time.[120] The Court's invalidation of New Deal legislation created "hostility toward judicial activism."[121] Additionally, the Court may have departed from previous Article V discussions in *Coleman* because it recognized that if the Court were to decide the case on the merits, it would be "controlling the very process employed to overrule the Court's interpretation of the Constitution."[122] As Millet has argued, "[a]llowing the least republican branch to review cases involving amendments intended to overrule that branch's opinions gives the Court the unchecked power to thwart those efforts."[123] Thus, judicial review of a convention's work product presents a genuine separation of powers issue, which the Court recognized in *Baker* as a reason for imposing the political question doctrine and other justiciability issues.[124]

While *Baker*'s illustration of political questions provides a bright line definition of justiciability, the Court's listed characteristics serve as descriptive factors rather than determinative criteria.[125] The Supreme

117. Baker, *supra* note 103, at 6.
118. Millet, *supra* note 103, at 748.
119. Baker, *supra* note 106, at 339–40.
120. *See* Millet, *supra* note 103, at 756.
121. *Id.*
122. *Id.*
123. *Id.* at 764.
124. Baker v. Carr, 369 U.S. 186, 210–11 (1962).
125. Millet, *supra* note 103, at 760–61. Millet writes:

> The Supreme Court's application of the political question doctrine since *Baker* demonstrates that the *Baker* criteria are not mechanical formulations, but simply descriptive of recurring characteristics which indicate a potential separation of powers issue. For example, application of the "textually demonstrable commitment" factor to the provision of article I, section 5, that "[e]ach House shall be the Judge of the Elections, Returns, and Qualifications of its own Members . . ." would certainly lead to the conclusion that controversies over the qualifications of members of Congress were committed to each House and, consequently, were non-justiciable. Yet, the Supreme Court held just the opposite in *Powell v. McCormack.*

Court's decision to grant certiorari to a case involving Article V will undoubtedly place the Court at the crossroads of its line of (albeit few) Article V cases and the political question doctrine. The Court's involvement could hinge on whether the conflict is over an amendment that "is intended to affect a prior Supreme Court constitutional interpretation."[126] Yet it has been said that "the Court retains for itself the power to frustrate the amendment effort and leaves the judicial review power unchecked"[127] if it injects itself into the amendment process. It would be entirely appropriate for the Court to rely on the political question doctrine and refrain from judgment,[128] Millet writes, as any ruling by the Court in that situation "would do violence to the notion of popular sovereignty, which, as noted by Alexander Hamilton, is the stream from which legitimate national power ought to flow."[129] To the extent that one adopts the position that an Article V convention exists to express the popular will when the ordinary process of government has ceased to do so, Supreme Court involvement might be wholly inappropriate, particularly if the Court purposes to protect its traditional prerogatives against such will.

C. Judicial Activism

Given the murky waters of the Supreme Court's role in any convention proceedings, would *any* involvement by the Court result in accusations of "judicial activism"? The Court has often discovered implied powers and doctrines in the Constitution nowhere to be found in its text—to wit, its arrogation in *Marbury v. Madison* of the very power of judicial review.[130] That case presented the seemingly simple question of the scope of the Court's original jurisdiction.[131] After answering the question at hand—whether Congress had the authority to amend the Court's original jurisdiction—the Court went further, stating "[i]t is emphatically the province and duty of the judicial department to say what the law is."[132] The Court went beyond the mere facts of the case to stymie Jefferson's

---

> While acknowledging the *Baker* factors, the Court in *Powell* held that the judiciary must first determine the meaning of the Constitution before ascertaining whether the Constitution committed a particular power to another branch. Through engaging in this inquiry, the Court, in essence, ruled upon the merits of the case.

Id. at 761–62.

126. *Id.* at 763.

127. *Id.*

128. *See id.* (citing Goldwater v. Carter, 444 U.S. 996 (1979) (Powell, J. concurring)).

129. *Id.* at 764 (citing THE FEDERALIST NO. 22 (Alexander Hamilton)).

130. 5 U.S. (1 Cranch) 137 (1803).

131. *Id.* at 146–47.

132. *Id.* at 177.

Democratic-Republicans—and exalt its own position—a textbook case of judicial activism.[133]

More recently, the Court's 2000 decision in *Bush v. Gore* was declared "one of the more extreme cases of judicial activism, made all the more extreme because it was issued by a group of justices who allegedly practice judicial restraint."[134] In *Bush v. Gore*, four of the nine justices thought that the Court should abstain from judgment in the case.[135] Justice Breyer's dissent, joined by Justices Stevens, Ginsburg, and Souter, illustrates this opposition:

> The Court was wrong to take this case. . . . The political implications of this case for the country are momentous. But the federal legal questions presented, with one exception, are insubstantial.[136]

The Court's decision in *Bush v. Gore* has caused the Court to "come under intense and withering criticism,"[137] but others have praised the Court for its decision, claiming the Court's "timely intervention was necessary to preserve the core foundations of the American democratic political system."[138] *Marbury v. Madison* and *Bush v. Gore* are but two decisions in more than 200 years of Supreme Court jurisprudence, but both serve as indications that the Court could be willing to take a case arising from a constitutional convention conflict even though the issue may present a nonjusticiable political question. However, just as in *Bush v. Gore*, the Court could lose a great deal of its institutional prestige and repute by even accepting the case, or, worse, by making the "wrong" decision. It is possible that "[i]f the [C]ourt's interpretation of the [C]onstitution w[as] sufficiently at odds with public opinion, [it] would lose . . . considerable prestige and the other branches of government would find a way to bypass the [C]ourt's decisions."[139] Of course, the quantity and tenor of any criticism will hinge on whose ox is being gored.

It is impossible to know for certain what role the Supreme Court would carve out for itself in the convention process until Congress actually calls a

---

133. *Id.* at 177; Dixon, *supra* note 16, at 931. The *Marbury* decision has been decried as "one of the most flagrant specimens of judicial activism . . . ." LEONARD W. LEVY, ORIGINAL INTENT AND THE FRAMERS' CONSTITUTION 75 (1988).

134. Joseph A. Reinert, *The Myth of Judicial Activism*, 29 VT. B.J. 35, 37 (2004) (citing Paula Alexander Becker & Richard J. Hunter, Jr., *A Review of the Supreme Court's 2000 Term: Is There a Consistent Theme?*, 38 HOUS. L. REV. 1463, 1465 (2002)).

135. Bush v. Gore, 531 U.S. 98, 144 (2000) (Breyer, J., dissenting).

136. *Id.*

137. Becker & Hunter, *supra* note 134, at 1466.

138. *Id.* at 1467 (citing RICHARD A. POSNER, BREAKING THE DEADLOCK: THE 2000 ELECTION, THE CONSTITUTION, AND THE COURTS ix (2001)).

139. George Mailath, Stephen Morris & Andrew Postlewaite, *Maintaining Authority* (Sept. 26, 2007) (unpublished paper), *available at* http://www.princeton.edu/%7Esmorris/pdfs/authority.pdf.

convention and conflict arises. Where the question is one interpreting the power granted by Article V's language, the Court has previously addressed similar issues with other provisions of the Constitution that grant powers to Congress.[140] In contrast, if the conflict is one related to an actual constitutional amendment, or if an amendment overrules a significant Court decision, the Court's intercession appears less likely.[141] In general, it seems most prudent to believe that the Court would likely follow its own precedent in *Baker v. Carr* and resolve an issue on the merits only if it presents a truly non-political, justiciable question.[142]

VI. THE ORGANIZATION OF AN ARTICLE V CONVENTION

This portion of the primer attempts to consolidate years of scholarly guesswork into one workable framework of practical procedures with respect to two specific issues involving an Article V convention: (1) how a convention should be organized, and (2) how it should operate. Before addressing questions regarding mode or operation of the convention once it is called and assembled, it is first necessary to discuss concerns about how the convention would initially come about, including how the convention would be called and who would comprise it.

A. The Calling of a Convention

Should Congress call a constitutional convention once it receives enough applications,[143] Congress may have the ability to control several aspects of a convention for practical and organizational reasons, including the convention's forum and how states are represented.

---

140. Congress's power to control interstate commerce is one example. *See generally* Wickard v. Filburn, 317 U.S. 111, 128 (1942) (noting that Congress's power to regulate interstate commerce includes the power to regulate prices of commodities and to regulate the practices affecting such prices); United States v. Darby, 312 U.S. 100, 114 (1941) (holding that the Fair Labor Standards Act of 1938 was within Congress's commerce power); Carter v. Carter Coal Co., 298 U.S. 238, 297 (1936) (noting that the commerce clause of the Constitution gives Congress the power to regulate interstate commerce); Gibbons v. Ogden, 22 U.S. (9 Wheat.) 1, 196 (1824) (noting that Congress has a vested power to regulate commerce). Additionally, the Supreme Court has interpreted the taxation and spending powers. *See generally* South Dakota v. Dole 483 U.S. 203, 206 (1987) (holding that Congress's indirect encouragement of uniformity of drinking ages among the states is a valid use of spending power); Steward Machine Co. v. Davis, 301 U.S. 548, 583 (1937) (holding that the tax imposed by Title IX of the Social Security Act of 1935 complies with the intent of the Constitution); United States v. Butler, 297 U.S. 1, 61 (1936) (holding that "[i]t does not follow that [when] the act is not an exertion of the taxing power and the exaction not a true tax, the statute is void . . . .").

141. *See supra* text accompanying notes 120–29.

142. *See* Baker v. Carr, 369 U.S. 186, 217 (1962).

143. U.S. CONST. art. V.

1. Forum Selection

Pursuant to its implied power under Article V, Congress can likely designate a time and place for the convention.[144] Although some scholars hold legitimate fears about Congress's role in convention proceedings,[145] these exclusively procedural questions seem to be an area that would fall under Congress's administrative authority. In addition to its administrative function, the federal government would probably take responsibility for funding the convention and compensating each delegate for his or her participation in the convention.

The most rational place to hold an Article V convention would be the nation's capital, Washington, D.C., where the delegates would have easy access to the full range of the federal government's resources. On the other hand, given current public opinion of Washington and the perceived disconnectedness between the federal government and the people, a convention, as a source of government expressly and intentionally alternative to Congress,[146] may be better off assembling in one of the nation's other metropolitan centers, such as Chicago, New York City, Los Angeles, or, in a nod to its historical roots, Philadelphia. However, Congress, as the body charged with determining the convention's meeting place,[147] would no doubt lean toward having the convention assemble in Washington.

2. State Representation

State representation and delegate selection are nuanced issues implicating significant national concerns. For example, two theories have been advanced with respect to how the states should be represented at a convention. The first holds that each state would have equal representation at a convention, regardless of population differences.[148] The second accounts for the states' unequal population and provides that each state would have representation at a convention proportionate to its number of

144. *See Proposing Amendments*, *supra* note 76, at 1075; *see also Convention Method*, *supra* note 113, at 1618 (stating that Congress's role in the convention process "should be limited to those 'housekeeping matters' which are necessary aspects of the implementation of Congress' [sic] duty to call a convention or its power to choose the mode of ratification of proposed amendments.").

145. Gunther, *supra* note 8, at 24 ("Congress seems to me to go well beyond legitimate bounds when it does more than setting up necessary machinery and when it goes on to impose substantive limitations on the scope and duration of convention deliberations.").

146. *See Convention Method*, *supra* note 113, at 1636 (noting that "the purpose of [this] method of amendment is to allow citizens to bypass an obdurate Congress in bringing about Constitutional change.").

147. *Proposing Amendments*, *supra* note 76, at 1075.

148. *See id.* at 1076.

representatives in Congress.[149] As previously discussed, the extent of Congress's power in this area is unclear, but "it would seem proper for Congress to determine which of a number of alternative methods of representation should be adopted."[150]After all, "[i]t would be difficult for the convention or the states to resolve this issue, since before a vote could be taken on the system of representation the states would have to decide the very issue which they were to decide."[151]

To answer questions regarding the uncertainties of a modern convention, some scholars suggest that we simply refer to "the historical precedent of the constitutional convention in Philadelphia in 1787."[152] If we followed the precedent established over two centuries ago, each state would be afforded an equal number of delegates at a convention.[153] However, the present differences in state populations would result in grossly malapportioned delegations.[154] Indeed, as Dixon wrote, "Although Article V may expressly authorize Senate malapportionment by exempting it from constitutional amendment, it does not authorize parallel convention malapportionment."[155] To illustrate the flaws of the equal state representation approach, imagine that a convention had been called to address the reapportionment issues of the mid-twentieth century. If the convention had adhered to equal state representation, the result would be paradoxical: a malapportioned convention attempting to address concerns stemming from malapportionment.[156]

This problem of malapportionment can be minimized by "modeling the national convention on the House of Representatives and basing it solely on congressional districts . . . ."[157] This method would limit the amount of delegates to 435 and allow the number of delegates from each state to be proportionate to its population, thereby creating a convention more representative of the American public.[158]

---

149. *See id.* at n.50.

150. *Id.* at 1076.

151. *Id.*

152. Brennan, *supra* note 9, at 69.

153. *Proposing Amendments*, *supra* note 76, at 1076.

154. *See* Dixon, *supra* note 16, at 945.

155. *Id.* But neither does it prohibit malapportionment. *See id.*

156. *See id.* (recognizing that it would be "an anomaly to establish a national convention to deal with state legislative apportionment which was itself malapportioned").

157. *Id.*

158. Another possibility would be to model the apportionment of electors on the Electoral College, whereby each state is allocated a number of presidential electors equal to the combined number of its Representatives and Senators. Of course, such a scheme would be subject to the same criticisms as the Electoral College. *See, e.g.*, Editorial, *Flunking the Electoral College*, N.Y. TIMES, Nov. 20, 2008, at A42.

3. Selection of Delegates

In addition to allocating the numbers of delegates per state, another critical issue is how the states would be represented at the convention, and uncertainties remain as to how the states would actually select their delegates. This is a crucial inquiry because the men and women who comprise the convention would be confronted simultaneously with the truly awesome responsibility of adequately confronting the concerns that prompted the states' petitions for change and the task of ensuring the ongoing prosperity of a document that has served as the beacon of republican government for over two centuries. As one commentator has noted, there is "no cause more worthy of any responsible citizen's best effort and total commitment" than serving as a convention delegate.[159]

It seems appropriate that the states, rather than Congress, have the power to determine how to choose their respective convention delegations.[160] If the states followed the precedent established by the Philadelphia convention, delegates would be nominated and appointed by the elected officials in the state legislatures.[161] This mechanism offers some benefits. For instance, state legislatures are likely most aware of those members of the state's citizenry who are capable of handling the responsibilities accompanying selection as a delegate. Moreover, as the bodies that petitioned Congress to call an amendatory convention, the state legislatures should have some authority over who represents their interests at that convention.

Despite these considerations, the assumption among most modern scholars is that convention delegates would be selected by popular election rather than legislative appointment.[162] While this system of popular election is consistent with the republican ideals that a convention represents, it too presents some significant concerns. For instance, after receiving campaign support from special interests in a contested and protracted election, some delegates, like some other elected officials, may feel obligated to side with those special interests rather than the interests of the delegate's constituents.[163]

Because each of these mechanisms presents benefits and drawbacks, perhaps a hybrid appointment-election system would be best. Under this system, a state legislature could select a number of the state's citizens as candidates for the delegate positions. The candidates would then be placed

---

159. *See* Brennan, *supra* note 9, at 71.

160. *Proposing Amendments*, *supra* note 76, at 1076 ("It would seem proper for each state to determine the procedure for the election of its delegates and the qualifications of the electors, since matters of this kind have traditionally been left to the states.").

161. *See* Brennan, *supra* note 9, at 69.

162. *See, e.g.*, Gunther, *supra* note 8, at 8 n.20 ("Popular election of delegates has been the assumption in most modern discussions of Article V constitutional conventions.").

163. *See* Gunther, *supra* note 8, at 8.

on a ballot to be voted on by the state's registered voters. Each of the citizens nominated as candidates would have the opportunity to campaign for a period of time prior to the general election, and the ballot box would ultimately determine which of these citizens would win the honor of representing his or her state at the convention.

B. The Operation of a Convention

Once the date and time has been selected and the delegates assembled, the convention would convene, albeit with serious questions about how to move forward. To date, no procedural framework has been established to guide convention delegates in their deliberations and proceedings.[164] Although some commentators suggest that the convention itself, rather than Congress, should have the power to "choose its own officers [and] adopt its own rules of procedures,"[165] a number of federal lawmakers have proposed guideline legislation in hopes of "remov[ing] some of the uncertainties about the convention route . . . ."[166]

Beginning with Senator Sam Ervin's efforts in the late 1960s and continuing into the latter half of the twentieth century through the work of Senators Jesse Helms and Orrin Hatch, several members of Congress have championed bills seeking to establish rules of procedure for calling, convening, and conducting a constitutional convention.[167] However, despite some limited success in the Senate,[168] "such legislation has never been enacted into law."[169] As a result, should a convention ever be called, delegates would be forced to spend significant time discussing, reviewing, and ultimately adopting procedural and voting rules before engaging in any substantive debate regarding the proposed changes to the Constitution.[170]

Although a convention would not be bound by any pre-existing procedural framework, it could draw guidance from the proposed legislation of the past few decades, especially with respect to major issues such as voting procedures.[171] The major questions that a convention would be forced to answer, in terms of voting, are whether matters would be decided on a one-vote-per-state or one-vote-per-delegate basis and whether

---

164. *Convention Method*, *supra* note 113, at 1612.

165. *Proposing Amendments*, *supra* note 76, at 1076.

166. Gunther, *supra* note 8, at 20.

167. *See* Lawrence Schlam, *State Constitutional Amending, Independent Interpretation, and Political Culture: A Case Study in Constitutional Stagnation*, 43 DEPAUL L. REV. 269, 344 n.276 (1994).

168. *Id.*

169. Paulsen, *supra* note 28, at 734 n.194.

170. Rogers, *supra* note 2, at 1015 ("The convention would . . . determine its own voting rules and procedures when the delegates from the States convene.").

171. *See generally* Schlam, *supra* note 167, at 344 n.276 (noting the past legislation that has been proposed to set procedures for the process of an Article V amendment).

a winning number of votes would be a simple majority or a two-thirds majority.[172]

First, the convention should consider whether to decide matters on a one-vote-per-state or one-vote-per-delegate basis.[173] The 1787 convention employed a unit voting approach in which each state cast only one vote.[174] However, the landscape of the nation has changed dramatically since 1787, and the unit voting approach of the first convention is contrary to the "one man, one vote" movement popularized in the 1960s and embodied by the Supreme Court's decisions in *Reynolds v. Sims* and its progeny.[175] The one-vote-per-delegate approach first proposed by Senator Ervin is more consistent with this trend.[176] Senator Ervin endorsed this approach out of deference to the "interests of majority rule" and recognized that under that framework, "the voting strength of each state [would] be in proportion to its population."[177] Furthermore, any preferences for the states voting as units would be appeased by the fact that "the states, as units, will have equal say in the ratification process."[178] Thus, a modern convention should use the one-vote-per-delegate approach as it ensures that each individual voice is heard on important matters of national interest.

Second, the convention must consider whether an amendment would be submitted to the states for ratification upon a simple majority vote or a two-thirds majority vote.[179] Although the text of Article V is silent on this issue, it does provide that in the method of congressional proposal, an amendment will not be submitted to the states for ratification until two-thirds of both houses of Congress vote in favor it.[180] Accordingly, "a two-thirds voting requirement [in the convention] would be analogous to the requirement for congressional proposal of amendments."[181] While it seems logical to apply this requirement to the convention method, proponents of a simple majority requirement have focused on the absence of such express language in the

---

172. *See* Sam J. Ervin, Jr., *Proposed Legislation to Implement the Convention Method of Amending the Constitution*, 66 MICH. L. REV. 875, 893 (1968).

173. *See generally id.* at 893 (discussing the advantages of a one-vote-per-delegate basis over a one-vote-per-state basis).

174. *See id.*

175. *See id.* at 876, 893; *see generally* Reynolds v. Sims, 377 U.S. 533 (1964) (holding that the Equal Protection Clause requires that seats in both houses of a bicameral legislature be apportioned substantially on a population basis).

176. *See* Ervin, *supra* note 172, at 893.

177. *Id.*

178. *Id.*

179. The approval of three-fourths of the delegates at the convention would not be without constitutional support, as amendments take effect only after approval of three-fourths of state legislatures. However, requiring three-fourths of the delegates to settle on a proposed amendment would demand a greater level of consensus than the two-thirds majority required for Congress to propose an amendment. *See* U.S. CONST. art. V.

180. U.S. CONST. art. V.

181. Ervin, *supra* note 172, at 893.

article itself.[182] For his part, Senator Ervin argued that a two-thirds "requirement would place an undue and unnecessary obstacle in the way of effective utilization of the convention amendment process."[183] Nevertheless, amending the Constitution is not meant to be easy,[184] and it makes sense to impose upon the convention delegates the same obstacles that face legislators: in this instance, a two-thirds majority requirement.

VII. CONCLUSION

Because an Article V constitutional convention has never occurred, its actual procedure remains mysterious.[185] However, by revisiting the proceedings of the original 1787 convention,[186] examining past instances where conventions have nearly been called,[187] analyzing draft proposals of legislation seeking to establish a framework for a convention,[188] and recognizing the current interests of our modern society,[189] we can continue to develop answers to many of the questions that lead some commentators to believe that a convention's risks outweigh its benefits.

First, Congress is obligated to call a convention to amend the Constitution upon the valid application of two-thirds of the states.[190] Furthermore, in calling the convention, Congress likely has the authority to set the dates, times, and place for the convention to occur.[191] Given malapportionment concerns, the number of delegates could be limited to 435 with each state represented in proportion to its level of representation in the House of Representatives.[192] Each state's citizenry could select delegates to represent their state at the convention[193]and once present, each delegate, consistent with the "one man, one vote" trend, could receive one vote in all voting procedures.[194] While the amount of congressional involvement in a convention is likely rather limited, the full extent of the Supreme Court's intervention in either the format of a convention or a resulting proposed amendment is more difficult to predict and may depend

182. *See id.*

183. *Id.*

184. Saikrishna B. Prakash & John C. Yoo, *The Origins of Judicial Review*, 70 U. CHI. L. REV. 887, 918 (2003).

185. Manne, *supra* note 9, at 135; *see also* Brennan, *supra* note 9, at 68.

186. *See* Ervin, *supra* note 172, at 893.

187. *See supra* Part II.A.

188. *See supra* notes 164–72 and accompanying text.

189. *See supra* notes 173–78 and accompanying text.

190. *See supra* Part IV.A.

191. *See Proposing Amendments*, *supra* note 76, at 1075.

192. *See* Dixon, *supra* note 16, at 945.

193. *See supra* Part VI.A.3.

194. *See supra* Part VI.B (discussing the merits of a one-vote-per-delegate system).

on the Court's composition.[195] However, it is clear that the President can play no formal part in the process.

Finally, the proper scope of the convention's deliberations will have to be addressed to guide Congress as it evaluates states' applications and decides what level of control it is constitutionally authorized to exert over the convention's deliberations and which amendments must be submitted to the states for approval.

195. *See supra* Part V.

PROPOSING CONSTITUTIONAL AMENDMENTS BY CONVENTION: RULES GOVERNING THE PROCESS

ROBERT G. NATELSON*

ABSTRACT

Much of the mystery surrounding the Constitution's state-application-and-convention amendment process is unnecessary: History and case law enable us to resolve most questions. This Article is the first in the legal literature to access the full Founding-Era record on the subject, including the practices of inter-colonial and interstate conventions held during the 1770s and 1780s. Relying on that record, together with post-Founding practices, understandings, and case law, this Article clarifies the rules governing applications and convention calls, and the roles of legislatures and conventions in the process. The goal of the Article is objective exposition rather than advocacy or special pleading.

ABSTRACT 693
I. INTRODUCTION 696
II. FOUNDING-ERA TERMINOLOGY 698
III. THE PURPOSE OF THE STATE-APPLICATION-AND-CONVENTION PROCESS 699
IV. THE ESSENCE OF ARTICLE V: GRANTS OF POWER TO DESIGNATED ASSEMBLIES 702
V. READING CONSTITUTIONAL GRANTS OF POWER: THE FIDUCIARY CONTEXT 703
A. The Centrality of Fiduciary Rules 703
B. The Doctrine of Incidental Authority 704
VI. OVERVIEW OF FOUNDING-ERA CONVENTIONS 706
VII. OTHER EVIDENCE—FOUNDING AND POST-FOUNDING 708
VIII. THE NATURE OF APPLICATIONS AND THE RULES GOVERNING THEM 709
A. The Nature of an Application 709
B. The Application Process is Not Subject to Normal Legislative Limitations, Such as Presentment to the Governor 710

---

* Robert G. Natelson, the author of *The Original Constitution: What It Actually Said and Meant*, is a constitutional historian. He is a Senior Fellow in Constitutional Jurisprudence at the Independence Institute in Golden, Colorado and at the Montana Policy Institute in Bozeman, Montana; Senior Fellow at the Goldwater Institute in Phoenix, Arizona; and until 2010 served as Professor of Law at the University of Montana. His works are listed at http://constitution.i2i.org/about.

C. *States May Rescind Applications* 712
D. *Applications Do Not Grow "Stale" with the Passage of Time* 712
IX. DEFINING THE SCOPE OF THE CONVENTION 715
A. *Founding-Era Convention Practice Before the 1787 Convention* 715
B. *Was the 1787 Federal Convention a "Runaway?"* 719
C. *Other Evidence that Applications Can Limit the Convention's Agenda* 723
X. THE CONVENTION CALL AND SELECTION OF DELEGATES 732
A. *Congress as a (Limited) Agent of the States* 732
B. *Congress's Role in Calling the Convention* 733
C. *Other Formalities in the Call* 737
D. *Enforcing the Duty to Call* 737
E. *The Composition of the Convention* 738
F. *Convention Discretion: The Rules* 740
G. *Convention Discretion: An Application May Not Limit the Convention to Specific Rules or Language* 742
H. *State Legislative Instructions* 747
XI. RULES GOVERNING TRANSMITTAL OF PROPOSALS TO THE STATES. 748
A. *What Happens if the Convention "Proposes" an Amendment Outside the Subject Assigned by the Applications?* 748
B. *Choosing a Mode of Ratification* 748
XII. CONCLUSION[1] 750

---

1. *Bibliographical Note*: This footnote collects alphabetically the secondary sources cited more than once in this Article. The listing reflects this article's reliance on several of the author's prior publications. The sources and short form citations used are as follows:

Charles L. Black, *Amending the Constitution: A Letter to a Congressman*, 82 YALE L.J. 189 (1972) [hereinafter Black, *Amending*].

RUSSELL L. CAPLAN, CONSTITUTIONAL BRINKMANSHIP: AMENDING THE CONSTITUTION BY NATIONAL CONVENTION (1988) [hereinafter CAPLAN].

Walter E. Dellinger, *The Recurring Question of the "Limited" Constitutional Convention*, 88 YALE L.J. 1623 (1979) [hereinafter Dellinger].

Ann Stuart Diamond, *A Convention for Proposing Amendments: The Constitution's Other Method*, 11 PUBLIUS: THE J. OF FEDERALISM 113 (1980) [hereinafter Diamond].

THE DOCUMENTARY HISTORY OF THE RATIFICATION OF THE CONSTITUTION (Merrill Jensen, John P. Kaminski, & Gaspare J. Saladino eds., 1976–2009) [hereinafter DOCUMENTARY HISTORY].

JONATHAN ELLIOT, 2, 3, 4 THE DEBATES IN THE SEVERAL STATE CONVENTIONS ON THE ADOPTION OF THE FEDERAL CONSTITUTION (2d ed. 1836) [hereinafter ELLIOT'S DEBATES].

Sam J. Ervin, Jr., *Proposed Legislation to Implement the Convention Method of Amending the Constitution*, 66 MICH. L. REV. 875 (1968) [hereinafter Ervin].

1, 2, 3 THE RECORDS OF THE FEDERAL CONVENTION OF 1787 (Max Farrand, ed., 1937) [hereinafter FARRAND'S RECORDS].

THE FEDERALIST NOS. 10, 40, 43 (James Madison), NOS. 82, 85 (Alexander Hamilton) (George W. Carey & James McClellan eds., 2001).

John M. Harmon, Constitutional Convention: Limitation of Power to Propose Amendments to the Constitution, 3 Op. O.L.C. 390 (1979) [hereinafter Harmon].

THE PUBLIC RECORDS OF THE STATE OF CONNECTICUT - FROM OCTOBER, 1776, TO FEBRUARY, 1778, INCLUSIVE (Charles J. Hoadly, ed., 1894) [hereinafter 1 HOADLY];

THE PUBLIC RECORDS OF THE STATE OF CONNECTICUT - FROM MAY, 1778, TO APRIL, 1780, INCLUSIVE (Charles J. Hoadly, ed., 1895) [hereinafter 2 HOADLY];

THE PUBLIC RECORDS OF THE STATE OF CONNECTICUT - FROM MAY, 1780, TO OCTOBER, 1781, INCLUSIVE (Charles J. Hoadly, ed., 1922) [hereinafter 3 HOADLY].

ROGER SHERMAN HOAR, CONSTITUTIONAL CONVENTIONS: THE NATURE, POWERS, AND LIMITATIONS (1917) [hereinafter HOAR].

JOURNALS OF THE CONTINENTAL CONGRESS 1774–1789 (Worthington Chauncy Ford, Gaillard Hunt, & Roscoe R. Hill eds., 1904–1936) [hereinafter JCC].

Paul G. Kauper, *The Alternative Amendment Process*, 66 MICH. L. REV. 903 (1968) [hereinafter Kauper].

GARY LAWSON, GEOFFREY P. MILLER, ROBERT G. NATELSON & GUY I. SEIDMAN, THE ORIGINS OF THE NECESSARY AND PROPER CLAUSE (2010) [hereinafter ORIGINS OF THE NECESSARY AND PROPER CLAUSE].

Robert G. Natelson, *Amending the Constitution by Convention: A More Complete View of the Founders' Plan* (2010), INDEPENDENCE INSTITUTE, http://constitution.i2i.org/files/2010/12/IP_7_2010_a.pdf [hereinafter Natelson, *Amending*].

Robert G. Natelson, *Judicial Review of Special Interest Spending: The General Welfare Clause and the Fiduciary Law of the Founders*, 11 TEX. REV. L. & POL. 239 (2007) [hereinafter Natelson, *Judicial Review*]

Robert G. Natelson, *Learning from Experience: How the States Used Article V Applications in America's First Century*, GOLDWATER INSTITUTE (Nov. 4, 2010), http://www.goldwaterinstitute.org/article/5353 [hereinafter Natelson, *First Century*].

I. INTRODUCTION

Article V of the United States Constitution allows either Congress or a "Convention for proposing Amendments" to propose formally constitutional amendments for ratification or rejection. The relevant language is as follows:

> The Congress, whenever two thirds of both Houses shall deem it necessary, shall propose Amendments to this Constitution, or, on the Application of the Legislatures of two thirds of the several States, shall call a *Convention for proposing Amendments*, which, in either Case, shall be valid to all Intents and Purposes, as Part of this Constitution, when ratified by the Legislatures of three fourths of the several States, or by Conventions in three fourths thereof, as the one or the other Mode of Ratification may be proposed by the Congress. . . .[2]

A convention for proposing amendments also has been called an *Article V convention*,[3] an *amendments convention,* and a *convention of the states*. As explained below, the common practice of referring to it as a "constitutional convention" or "con-con," is inaccurate.[4]

When two thirds of the state legislatures apply to Congress for a convention for proposing amendments, the Constitution requires Congress to call one.[5] Throughout this paper, this procedure is referred to as the *state-application-and-convention process*. The Framers inserted the procedure primarily to enable the people, through their state legislatures, to make changes in the Constitution without the consent of Congress.[6] The Framers' purpose, explained to the ratifying public as such, was to enable the people

---

ROBERT G. NATELSON, THE ORIGINAL CONSTITUTION: WHAT IT REALLY SAID AND MEANT (2010) [hereinafter NATELSON, ORIGINAL CONSTITUTION].

Robert G. Natelson, *Tempering the Commerce Power*, 68 MONT. L. REV. 95 (2007) [hereinafter Natelson, *Tempering*].

Note, *Proposing Amendments to the United States Constitution by Convention*, 70 HARV. L. REV. 1067 (1957) [hereinafter Note, *Amendments*].

Michael Stokes Paulsen, *A General Theory of Article V: The Constitutional Lessons of the Twenty-seventh Amendment*, 103 YALE L.J. 677 (1993) [hereinafter Paulsen].

Grover Joseph Rees III, *The Amendment Process and Limited Constitutional Conventions*, 2 BENCHMARK 66 (1986) [hereinafter Rees, *Amendment Process*].

2. U.S. CONST. art. V (emphasis added).

3. Although strictly speaking state ratifying conventions also are "Article V conventions."

4. *See infra* Part IX.A.

5. *See infra* Part X.B.

6. *See infra* Part III.

to restrain Congress if it should exceed or abuse its powers, or if the people wished to reduce congressional authority.[7] In a sense, the state-application-and-convention process is the federal analogue of the state voter initiative, whereby the electorate can bypass the legislature by adopting laws or amending the state constitution.[8]

Although the state-application-and-convention process has never been carried to completion, there have been many application campaigns.[9] Some failed only because Congress responded by proposing the sought-for amendments.[10] Others enjoyed insufficient popular support.[11] In recent years, however, such campaigns have been discouraged because of uncertainty about the legal rules governing the state-application-and-convention process—uncertainty promoted by persons and groups both on the political left and political right.[12]

Most of that uncertainty is needless, the product of alarmism and lack of knowledge. I wrote this paper in the belief that, whatever the merits of the process, light is better than darkness. To answer central questions, I rely on the constitutional text, judicial decisions,[13] application practice over the

---

7. *See infra* notes 20–21 and accompanying text.

8. *See, e.g.*, Rees, *Amendment Process*, *supra* note 1, at 83 (describing the process as "the closest thing the Constitution provides to the opportunity for a national referendum").

9. *See generally* CAPLAN, *supra* note 1, at 36–89 (describing campaigns through the 1980s); Natelson, *First Century*, *supra* note 1 (describing campaigns from 1789 through 1913).

10. *See generally* Natelson, *First Century*, *supra* note 1.

11. *Id.*

12. CAPLAN, *supra* note 1, at vii–viii, 146–47 (quoting various public figures, mostly on the political left); Art Thompson, *Help Stop the New Drive for a Constitutional Convention*, YOUTUBE (Oct. 18, 2010), http://www.youtube.com/watch?v=ggepQ6DtjP4 (presenting a video message from Art Thompson, president of the deeply conservative John Birch Society).

13. At one time, some argued that the courts should take no jurisdiction over Article V issues—that Congress, not the judiciary, should referee the process. Article V issues were said to be "political questions" inappropriate for judicial resolution. Coleman v. Miller, 307 U.S. 433, 450 (1939) (supporting the view from a four-justice concurring opinion and a brief *dictum* from the majority). However, *Coleman* has come under very heavy criticism, see, e.g., Rees, *Amendment Process*, *supra* note 1, at 98–107, and has not been followed. One scholar has accurately described the case as an "aberration." Walter Dellinger, *The Legitimacy of Constitutional Change: Rethinking the Amendment Process*, 97 HARV. L. REV. 386, 389 (1983). Today, the courts consciously reject the "hands-off" rule of the dictum and concurrence. *E.g.*, Dyer v. Blair, 390 F. Supp. 1291, 1301 (N.D. Ill. 1975) (explicitly rejecting, in a decision by the future Justice Stevens, the "political question" portion of *Coleman*); AFL-CIO v. Eu, 686 P.2d 609 (Cal. 1984) (declining to follow the "political question" doctrine from *Coleman*); *see also* Carmen v. Idaho, 459 U.S. 809 (1982), *vacating as moot* Idaho v. Freeman, 529 F. Supp. 1107, 1155 (D. Idaho 1981); Kimble v. Swackhamer, 439 U.S. 1385, 1387–88 (1978) (Rehnquist, J., sitting as a circuit judge, upholding Nevada's use of non-binding referenda on pending constitutional amendments).

Rejection of *Coleman* is implicit in *Powell v. McCormick*. 395 U.S. 486 (1969)

past two centuries, some insights from other scholars,[14] and a more thorough examination of relevant Founding-Era sources than previously has appeared in the legal literature.

Unlike most law review articles, this paper is not designed to be a work of advocacy. It was not written to advance any agenda other than the dissemination of knowledge about a little-understood part of our Constitution. When the evidence conflicted with my wishes or required me to revise my views, I followed the evidence wherever it led.

II. FOUNDING-ERA TERMINOLOGY

In discussing the Founding Era, I refer to several different groups of people.[15] The *Framers* were the fifty-five men who drafted the Constitution at the federal convention in Philadelphia, between May 29, 1787 and September 17, 1787. The *Ratifiers* were the 1,648 delegates at the thirteen state ratifying conventions held from November, 1787 through May 29, 1790. The *Federalists* were those participants in the public ratification debates who argued for adopting the Constitution. Their opponents were *Anti-Federalists*.

In this paper, the term *Founders* includes all who played significant roles in the constitutional process, whether Framers, Ratifiers, Federalists, or Anti-Federalists. Also among the Founders were the members of the Confederation Congress, 1781–89, and the members of the initial session of the First Federal Congress, 1789. Many Founders fit into more than one category. For example, James Madison was a Framer, Ratifier, and a leading Federalist, while Elbridge Gerry was a Framer and Anti-Federalist, but not a Ratifier.[16]

As used in this paper, the *original understanding* is the Ratifiers' subjective understanding, to the extent recoverable, of a provision in the Constitution—i.e., what those who voted for ratification actually understood the Constitution to mean. The *original meaning*, often called "original public meaning," is the objective meaning of a provision to a

---

(refusing to apply the political question doctrine when ruling directly against Congress). Although the judiciary has applied the "political question" doctrine to some Article V cases, in each of those cases, special facts called for abstention.

Thus, there is no general principle that Article V issues are not justiciable. On the contrary, a respectably long series of court rulings on Article V extends from 1798 to modern times. *See infra passim.*

14. Unfortunately, good scholarship on this subject is rare; most of the writing is poorly-researched, agenda-driven, or both. *See* Natelson, *Amending*, *supra* note 1, at 4, and accompanying notes.

15. *See* NATELSON, ORIGINAL CONSTITUTION, *supra* note 1, at 9–11.

16. *See generally* 1–2 FARRAND'S RECORDS, *supra* note 1 (detailing involvement of individuals throughout the process).

reasonable person at the time. *Original intent* is the subjective view of the Framers, to the extent recoverable.

Under Founding-Era jurisprudence, legal documents were interpreted according to the "intent of the makers," if available, and otherwise by the original meaning.[17] In the case of a constitution, the "intent of the makers" was the original understanding of the Ratifiers. Original intent did not have independent legal significance, but could serve as evidence of original understanding and original meaning.[18]

III. THE PURPOSE OF THE STATE-APPLICATION-AND-CONVENTION PROCESS

The Founding-Era record tells us that the two procedures for proposing amendments were designed to be equally usable, valid, and effective.[19] Congress received power to initiate amendments because the Framers believed that Congress's position would enable it readily to see defects in the system.[20] However, Congress might become abusive or refuse to adopt a necessary or desirable amendment—particularly one to curb its own power.[21] As one Anti-Federalist writer predicted, "[W]e shall never find

---

17. *See generally* Robert G. Natelson, *The Founders' Hermeneutic: The Real Original Understanding of Original Intent*, 68 OHIO ST. L.J. 1239 (2007). Professor Richard S. Kay concludes that *The Founder's Hermeneutic* "more or less settles the case to the contrary" of the widespread belief that Founding-Era interpreters relied only on original meaning and did not consider subjective understanding. Richard S. Kay, *Original Intention and Public Meaning in Constitutional Interpretation*, 103 NW. U. L. REV. 703, 709 (2009).

18. *See generally* Natelson, *supra* note 17.

19. *See infra* Part III; *see also* Diamond, *supra* note 1, at 114, 125 (emphasizing that the two methods were to be alternative means to the same end); *Letters from the Federal Farmer to the Republican*, Letters IV–V, Oct. 12, 1787, *reprinted in* 19 DOCUMENTARY HISTORY, *supra* note 1, at 231, 237, 239 (2003) ("No measures can be taken towards amendments, unless two-thirds of the congress, or two-thirds of the legislatures of the several states shall agree."); *cf.* Ervin, *supra* note 1, at 882 ("It is clear that neither of the two methods of amendment was expected by the Framers to be superior to the other or easier of accomplishment.").

20. 2 FARRAND'S RECORDS, *supra* note 1, at 558 (Sept 10, 1787) (Madison paraphrasing Alexander Hamilton as stating, "The National Legislature will be the first to perceive and will be most sensible to the necessity of amendments . . . .").

21. 1 FARRAND'S RECORDS, *supra* note 1, at 202–03 (Jun. 11, 1787), paraphrasing George Mason in discussing a resolution "for amending the national Constitution hereafter without consent of Natl. Legislature" as follows:

> Amendments therefore will be necessary, and it will be better to provide for them, in an easy, regular and Constitutional way than to trust to chance and violence. It would be improper to require the consent of the Natl. Legislature, because they may abuse their power, and refuse their consent on that very account. The opportunity for such an abuse, may be the fault of the Constitution calling for amendmt.

two thirds of a Congress voting or proposing anything which shall derogate from their own authority and importance."[22] In that eventuality, the state-application-and-convention procedure would permit the state legislatures to take corrective action.[23]

In the New York legislature, Samuel Jones explained the plan this way:

> The reason why there are two modes of obtaining amendments prescribed by the constitution I suppose to be this—it could not be known to the framers of the constitution, whether there was too much power given by it or too little; they therefore prescribed a mode by which Congress might procure more, if in the operation of the government it was found necessary; and they prescribed for the states a mode of restraining the

---

Mason was supported on this point by Edmund Randolph. *Id.* Ratification discussions in New York also contemplated a method of amendment separate from the national legislature:

> The amendments contended for as necessary to be made, are of such a nature, as will tend to limit and abridge a number of the powers of the government. And is it probable, that those who enjoy these powers will be so likely to surrender them after they have them in possession, as to consent to have them restricted in the act of granting them? Common sense says—they will not.

A PLEBEIAN, AN ADDRESS TO THE PEOPLE OF THE STATE OF NEW YORK (1788), *reprinted in* 20 DOCUMENTARY HISTORY *supra* note 1, at 942, 944 (2004).

22. Letter from An Old Whig, PHILA. INDEP. GAZETTEER (Oct. 12, 1787), *reprinted in* 13 DOCUMENTARY HISTORY, *supra* note 1, at 376, 377 (1981).

23. 3 ELLIOT'S DEBATES, *supra* note 1, at 101, quoting George Nicholas at the Virginia ratifying convention:

> [Patrick Henry] thinks amendments can never be obtained, because so great a number is required to concur. Had it rested solely with Congress, there might have been danger. The committee will see that there is another mode provided, besides that which originated with Congress. On the application of the legislatures of two thirds of the several states, a convention is to be called to propose amendments.

See also 4 ELLIOT'S DEBATES, at 177 (James Iredell, at the North Carolina ratifying convention):

> The proposition for amendments may arise from Congress itself, when two thirds of both houses shall deem it necessary. If they should not, and yet amendments be generally wished for by the people, two thirds of the legislatures of the different states may require a general convention for the purpose, in which case Congress are under the necessity of convening one. Any amendments which either Congress shall propose, or which shall be proposed by such general convention, are afterwards to be submitted to the legislatures of the different states, or conventions called for that purpose, as Congress shall think proper, and, upon the ratification of three fourths of the states, will become a part of the Constitution.

> powers of the government, if upon trial it should be found they had given too much.[24]

With his customary vigor, the widely-read Federalist essayist Tench Coxe, then serving in the Confederation Congress, described the role of the state-application-and-convention procedure:

> It has been asserted, that the new constitution, when ratified, would be fixed and permanent, and that no alterations or amendments, should those proposed appear on consideration ever so salutary, could afterwards be obtained. A candid consideration of the constitution will show this to be a groundless remark. It is provided, in the clearest words, that Congress shall be *obliged* to call a convention on the application of two thirds of the legislatures; and all amendments proposed by such convention, are to be *valid* when approved by the conventions or legislatures of three fourths of the states. It must therefore be evident to every candid man, that two thirds of the states can *always* procure a general convention for the purpose of amending the constitution, and that three fourths of them can introduce those amendments into the constitution, although the President, Senate and Federal House of Representatives, should be *unanimously* opposed to each and all of them. Congress therefore cannot hold *any power,* which three fourths of the states shall not approve, on *experience.*[25]

Madison stated it more mildly in *Federalist No. 43*: The Constitution "equally enables the General, and the State Governments, to originate the

---

24. NEW YORK ASSEMBLY DEBATES (Feb. 4, 1789), *in* 23 DOCUMENTARY HISTORY, *supra* note 1, at 2523–24 (2009). During the same debate, John Lansing, Jr., a former delegate to the federal convention, gave additional reasons for the alternative routes to amendment:

> In the one instance we submit the propriety of making amendments to men who are sent, some of them for six years, from home, and who lose that knowledge of the wishes of the people by absence, which men more recently from them, in case of a convention, would naturally possess. Besides, the Congress, if they propose amendments, can only communicate their reasons to their constituents by letter, while if the amendments are made by men sent for the express purpose, when they return from the convention, they can detail more satisfactorily, and explicitly the reasons that operated in favour of such and such amendments—and the people will be able to enter into the views of the convention, and better understand the propriety of acceding to their proposition.

Id. at 2523.

25. "A Friend of Society and Liberty," PA. GAZETTE, Jul. 23, 1788, *reprinted in* 18 DOCUMENTARY HISTORY, *supra* note 1, at 277, 283–84 (1995). Coxe made the same points in *A Pennsylvanian to the New York Convention*, PA. GAZETTE, Jun. 11, 1788, *reprinted in* 20 DOCUMENTARY HISTORY, *supra* note 1, at 1139, 1142 (2004). Coxe was Pennsylvania's delegate to the Annapolis convention.

amendment of errors, as they may be pointed out by the experience on one side or on the other."[26]

Thus, the state-application-and-convention process was inserted as a way for the people to amend the Constitution through the state legislatures, bypassing Congress.

IV. THE ESSENCE OF ARTICLE V: GRANTS OF POWER TO DESIGNATED ASSEMBLIES

Article V envisions roles in the amendment process for four distinct sorts of gatherings, groups that I sometimes refer to in this paper as *Article V assemblies.* The four are Congress, state legislatures, state ratifying conventions, and conventions for proposing amendments. Article V grants eight distinct enumerated powers to these assemblies—four at the *proposal* stage and four at the *ratification* stage. At the proposal stage, Article V:

> (1) grants to two thirds of each house of Congress authority to "propose" amendments,
> (2) grants to two thirds of the state legislatures power to make "Application" for a convention for proposing amendments,
> (3) grants to Congress power to "call" that convention, and
> (4) grants to the convention authority "for proposing" amendments.[27]

At the ratification stage, Article V:

> (1) authorizes Congress to "propose" whether ratification shall be by state legislatures or state conventions;
> (2) if Congress selects the former method, authorizes three fourths of state legislatures to ratify;
> (3) if Congress selects the latter method, impliedly empowers, and requires, each state to call a ratifying convention; and
> (4) empowers three fourths of those conventions to ratify.[28]

---

26. THE FEDERALIST NO. 43, *supra* note 1, at 228. Similarly, at the North Carolina ratifying convention, the following colloquy took place:

> Mr. BASS observed, that it was plain that the introduction of amendments depended altogether on Congress.
> Mr. IREDELL replied, that it was very evident that it did not depend on the will of Congress; for that the legislatures of two thirds of the states were authorized to make application for calling a convention to propose amendments, and, on such application, it is provided that Congress *shall* call such convention, so that they will have no option.

4 ELLIOT'S DEBATES, *supra* note 1, at 178.

27. U.S. CONST. art. V.

28. *Id.*

When an Article V assembly exercises an Article V action, it performs, in the phrase of the Supreme Court, a "federal function."[29] Thus, a state convention ratifying an amendment, and a state legislature either applying for a convention or ratifying an amendment, act under the appropriate Article V grant, rather than pursuant to powers reserved in the state.[30] Similarly, under Article V Congress does not perform as the federal legislature, but as an assenting body.

V. READING CONSTITUTIONAL GRANTS OF POWER: THE FIDUCIARY CONTEXT

A. The Centrality of Fiduciary Rules

Central to understanding the Constitution's power-grants, including those in Article V, is first to understand that the Founders assumed those grants would be subject to the rules imposed on private fiduciaries.

In Founding-Era political theory, legitimate government was, in John Locke's phrase, a "fiduciary trust."[31] For this reason, the Founders frequently described public officials by fiduciary names, such as "trustees" and "agents."[32] The Founders did not see the public trust standard as merely an ideal but as a core principle of public law.[33] This principle was to be enforced in several ways, including the traditional remedy for violation of

---

29. Leser v. Garnett, 258 U.S. 130, 137 (1922); *see also* Opinion of the Justices to the Senate, 366 N.E.2d 1226 (Mass. 1977); State *ex rel.* Donnelly v. Myers, 186 N.E. 918 (Ohio 1933); *In re* Opinion of the Justices, 172 S.E. 474 (N.C. 1933); Prior v. Norland, 188 P. 727 (Colo. 1920).

30. United States v. Sprague, 282 U.S. 716, 733 (1931); Hawke v. Smith, 253 U.S. 221 (1920); Dyer v. Blair, 390 F. Supp. 1291, 1308 (N.D. Ill. 1975) (Stevens, J.) ("The delegation [from Article V] is not to the states but rather to the designated ratifying bodies."); *cf. Sprague*, 282 U.S. 716 (Article V as a grant to Congress *qua* Congress, not to the U.S. government).

31. I have written extensively on this subject, and my conclusions have not been contested by other scholars. *See generally* ORIGINS OF THE NECESSARY AND PROPER CLAUSE, *supra* note 1, at 52–60; NATELSON, ORIGINAL CONSTITUTION, *supra* note 1, at 23–25 (discussing the Founders' view of public trust, the powers of agents, and the role of impeachment); Natelson, *Judicial Review*, *supra* note 1 (describing the general content of eighteenth-century fiduciary law); Robert G. Natelson, *The Agency Law Origins of the Necessary and Proper Clause*, 55 CASE W. RES. L. REV. 243 (2004) (discussing the powers of agents under eighteenth-century law); and Robert G. Natelson, *The Constitution and the Public Trust*, 52 BUFF. L. REV. 1077 (2004) [hereinafter Natelson, *The Constitution and the Public Trust*] (documenting the Founders' belief in fiduciary government).

32. *See* Natelson, *Judicial Review*, supra note 1, at 246; Natelson, *The Constitution and the Public Trust*, *supra* note 31, at 1084.

33. *See* Natelson, *The Constitution and the Public Trust*, *supra* note 31, at 1088 (discussing "the role of the public trust doctrine in drafting, submission, and ratification of the Constitution").

public sector fiduciary duty (or, as it usually was called, "breach of trust")—that is, impeachment-and-removal.[34]

During the framing and ratification process, participants frequently assessed issues according to fiduciary standards.[35] Thus, people discussed whether the delegates to the federal convention had exceeded their authority, whether the Constitution would promote fiduciary government, and whether other options might better serve that purpose.[36]

Eighteenth-century fiduciary law differed somewhat from modern law in its terminology and classifications, but the underlying principles were much the same. Three rules are particularly important for our purposes:

> (1) The wording of the instrument by which the principal empowered the fiduciary, read in light of its purposes, defined the scope of the latter's authority.[37]
> (2) A fiduciary was required to remain within the scope of this authority.[38] Of course, this rule did not prevent the fiduciary from *recommending* the action to the principal. However, this recommendation had no legal force unless adopted by the principal.[39]
> (3) If under the same instrument a fiduciary served more than one person, the fiduciary was required to treat them all equally and fairly—or, in the language of the law, "impartially."[40]

B. The Doctrine of Incidental Authority

In absence of agreement to the contrary, the scope of a fiduciary's authority included not only powers granted in words ("express" or "principal" powers), but also power "incidental" thereto.[41] This concept, and the rules by which incidental powers were defined, comprised the legal doctrine of *incidental authority*. The doctrine assured that a fiduciary received sufficient capacity to carry out the intent or purpose behind the grant.[42] Unlike the Articles of Confederation,[43] the Constitution incorporated the doctrine of incidental authority.

---

34. NATELSON, ORIGINAL CONSTITUTION, *supra* note 1, at 203–07.

35. *See* Natelson, *The Constitution and the Public Trust*, *supra* note 31, at 1136.

36. *See* sources cited *supra* note 31.

37. Natelson, *Judicial Review*, *supra* note 1, at 256.

38. *Id.* at 255–57.

39. Natelson, *Amending*, *supra* note 1, at 6.

40. *Id.* at 262–67.

41. This subject is fully developed in ORIGINS OF THE NECESSARY AND PROPER CLAUSE, *supra* note 1, at 60–68, 80–83.

42. *Id.* at 82–83.

43. Article II of the Articles of Confederation excluded the doctrine of incidental authority by this language: "Each State retains its sovereignty, freedom and independence, and every power, jurisdiction and right, which is not by this confederation *expressly* delegated to the United States, in Congress assembled." ARTICLES OF CONFEDERATION of

By the time of the Founding, that doctrine was a well-developed and prominent component of Anglo-American jurisprudence.[44] Under its rules, for Power B to be incidental to Power A, several requirements had to be met. First, Power B had to be less valuable and less important—that is, subsidiary—to Power A. This often was expressed by saying that a principal power had to be more "worthy" than its incident.[45] Hence, a document entrusting a bailiff with management of an estate generally included incidental authority to make leases at will, but not to lease for a term.[46] Moreover, Power B had to be either *customary* for exercising Power A or so *necessary* to the exercise of Power A that the agent's work would be subject to "great prejudice"[47] unless Power B were included.[48] But neither custom nor "great prejudice" was sufficient; subsidiarity was required as well.[49]

The Necessary and Proper Clause expressly acknowledged the grant of incidental powers to Congress.[50] In fact, the word "necessary" was a legal term of art meaning "incidental."[51] However, as leading Federalists explained during the ratification debates, the Clause actually bestowed no authority. Rather, it was an acknowledgment or recital[52] that the Constitution—like most other power-granting documents, but unlike the Articles of Confederation—incorporated the incidental authority doctrine. The doctrine would have applied even in absence of the Clause.[53]

Incidental authority, therefore, accompanies not only congressional powers, but all other powers granted by the Constitution. For example, Article II, which lists the President's powers, includes no "necessary and

---

1781, art. II (emphasis added).

44. ORIGINS OF THE NECESSARY AND PROPER CLAUSE, *supra* note 1, at 60.

45. *Id.* at 61–62.

46. *Id.* at 65.

47. *Id.* at 65.

48. *Id.* at 64–66.

49. Moreover, as Chief Justice Marshall pointed out, the real goal for exercising the incidental power had to be furtherance of the principal. An incidental power could not be exercised for its own sake on the "pretext" of exercising the principal. McCulloch v. Maryland, 17 U.S. (4 Wheat.) 316, 423 (1819). Today, Congress frequently regulates activities "substantially affect[ing]" interstate commerce so as to govern those activities, not because doing so is necessary or customary to regulating commerce. Natelson, *Tempering*, *supra* note 1, at 122–24.

50. U.S. CONST. art. I, § 8, cl. 18 ("The Congress shall have Power . . . . To make all Laws which shall be necessary and proper for carrying into Execution the foregoing Powers, and all other Powers vested by this Constitution in the Government of the United States or in any Department or Officer thereof. ").

51. ORIGINS OF THE NECESSARY AND PROPER CLAUSE, supra note 1, at 64.

52. *See id.* at 97–108.

53. *See also* Natelson, *Tempering*, *supra* note 1, at 101–02 (explaining that Chief Justice John Marshall, who wrote the opinion in *McCulloch*, the greatest of Necessary and Proper Clause cases, fully agreed).

proper" language, but the President enjoys incidental authority.[54] Similarly, the grants in Article V to conventions and state legislatures[55] carry incidental powers with them.[56]

What is the scope of those incidents? The answer to that rests largely in Founding-Era custom[57]—specifically the convention practices of the time. As the next Part shows, conventions were common enough for their practices to have become standardized.

VI. OVERVIEW OF FOUNDING-ERA CONVENTIONS

The founding generation understood a political "convention" to be an assembly, other than a legislature, designed to serve an ad hoc governmental function.[58] The British brought about regime changes in 1660 and 1689 through "convention Parliaments."[59] During the latter year, the American colonists held at least four conventions of their own.[60] The colonists continued to resort to the device over the ensuing decades.[61]

---

54. *See* U.S. CONST. art. II. The famous debate in the First Congress over whether the President could remove federal officers without senatorial consent was won by those who claimed that the power to remove was incidental either to the power to appoint or to the executive power generally. The debate is found at 1 ANNALS OF CONG. 473-608 (1789) (Joseph Gales ed., 1834), *available at* http://international.loc.gov/cgi-bin/ampage?collId=llac&fileName=001/llac001.db&recNum=51. Note that the debate and resolution occurred while the ratifications of two states, North Carolina and Rhode Island, were still in doubt.

55. United States v. Sprague, 282 U.S. 716, 733 (1931) ("The fifth article does not purport to delegate any governmental power to the United States . . . . On the contrary . . . that article is a grant of authority by the people to Congress, and not to the United States.").

56. The Necessary and Proper Clause does not apply because that Clause applies only to the "Government of the United States" and "Department[s] or Officer[s] thereof." U.S. CONST. art. I, § 8, cl. 18.

At a conference at Cooley Law School on September 16, 2010, a participant cited *Sprague* for the proposition that Article V was not open to construction, and so granted no incidental powers. *See Cooley Article V Symposium*, 28 COOLEY L. REV. (forthcoming Summer 2011). The presentations of various speakers at this symposium are available on YouTube. *See generally* http://www.youtube.com (In query field, search for "Cooley Article V Symposium"). However, *Sprague* involved not the entirety of Article V, but only unambiguous language where no construction or supplementation was necessary. *Sprague*, 282 U.S. at 732.

57. *See infra* Part VI.

58. Natelson, *Amending*, *supra* note 1, at 6; *see also In Re* Opinion of the Justices, 167 A. 176, 179 (Me. 1933) ("The principal distinction between a convention and a Legislature is that the former is called for a specific purpose, the latter for general purposes.").

59. Natelson, *Amending*, *supra* note 1, at 6.

60. CAPLAN, *supra* note 1, at 5–6 (discussing two conventions in Massachusetts, one in New York, and one in Maryland).

61. *Id.* at 7–9.

During the Founding Era it became one of their favorite methods of solving political problems.[62]

Many Founding-Era conventions were single-polity affairs, held within a colony or state, with delegates representing the people directly.[63] Others were interstate or, as they came to be called, "federal."[64] The initial interstate convention of the Founding Era was the First Continental Congress (1774), which despite being denoted a "Congress,"[65] qualified as a convention and was understood to be one.[66] There were at least ten other interstate conventions held after the Declaration of Independence and before the meeting of the Constitutional Convention in 1787: two in Providence, Rhode Island (1776-77 and 1781); one in Springfield, Massachusetts (1777); one in York, Pennsylvania (1777);[67] one in New Haven, Connecticut (1778); two in Hartford, Connecticut (1779 and 1780); one in Philadelphia (1780), one in Boston (1780), and one in Annapolis (1786).[68] Attendance at Founding-Era conventions ranged from three states

---

62. Natelson, *Amending*, *supra* note 1, at 6.

63. HOAR, *supra* note 1, at 2–10 (describing state constitutional conventions at the Founding); *see also* CAPLAN, *supra* note 1, at 8–16 (discussing conventions); *cf. Opinion of the Justices*, 167 A. at 179 (noting that conventions within states directly represented the people).

64. Natelson, *Amending*, *supra* note 1, at 6, 11.

65. The term "congress" commonly denoted a meeting of sovereignties. *See, e.g.*, THOMAS SHERIDAN, A COMPLETE DICTIONARY OF THE ENGLISH LANGUAGE (2d ed. 1789) (unpaginated) (defining "congress" in part as "an appointed meeting for settlement of affairs between different nations").

66. *E.g.*, 1 JCC, *supra* note 1, at 17 (1904) (quoting the credentials of the Connecticut delegates, empowering them to attend the "congress, or convention of commissioners, or committees of the several Colonies in British America"). The Second Continental Congress (1775–1781) arguably also was a convention, but because it acted as a regular government for more than six years, I have not treated it as such. The Confederation Congress (1781–1789) was a regularly established government.

67. On the York Convention, see *infra* note 159 and accompanying text.

68. For a summary of special purpose conventions, *see* CAPLAN , *supra* note 1, at 17–21, 96. Caplan mentions the Boston Convention, which is also referenced at 17 JCC, *supra* note 1, at 790 (1910) (Aug. 29, 1780) and 18 JCC, *supra* note 1, at 932 (1910) (Oct. 16, 1780). The journals of the conventions are reproduced in: 1 HOADLY, *supra* note 1, at 585–620 (reproducing journals from the Providence Convention (Dec. 25, 1776 to Jan. 3, 1777), the Springfield Convention, and the New Haven Convention); 2 HOADLY, *supra* note 1, at 562–79 (reproducing journals from the Hartford Convention (Oct. 1779) and the Philadelphia Convention (Jan. 1780)); 3 HOADLY, *supra* note 1, at 559–76 (reproducing journals from the Boston Convention, the Harford Convention (Nov. 1780), and the Providence Convention (June 1781)). The roster and recommendations of the Annapolis Convention may be found at *Proceedings of Commissioners to Remedy Defects of the Federal Government*, THE AVALON PROJECT: DOCUMENTS IN LAW, HISTORY AND DIPLOMACY, *available at* http://avalon.law.yale.edu/18th_century/annapoli.asp.

to twelve.[69] On their rosters one sees certain names repeatedly—enough to promote crystallization of common practices.[70]

Each interstate convention was called by state legislatures, sometimes pursuant to congressional recommendation.[71] They were modeled on conventions attended by international diplomats,[72] and consisted of delegates serving as agents for their respective state legislatures. The delegates were empowered by documents called "commissions" or "credentials," and, like other agents, were bound by the scope of their authority.[73] They were subject to additional legislative instructions.[74] Each state delegation formed a unit, often called a "committee."[75] The gathering as a whole sometimes was referred to a convention of "the states"[76] or a convention of "committees."[77]

As a result of all this experience, federal convention customs, practices, and protocols were fairly well standardized when Article V was written. In the ensuing pages, I shall cite those customs, practices, and protocols as relevant issues arise.

VII. OTHER EVIDENCE—FOUNDING AND POST-FOUNDING

Many other sources offer insight into the state-application-and-convention process. Information on the original meaning of Article V comes from eighteenth-century dictionaries, debates over the Constitution, material from the first session of the First Congress, including the first two

---

69. *See* sources cited *supra* note 68.

70. *See* sources cited *supra* note 68 (listing among their attendees such Constitutional Convention delegates as John Dickinson, Alexander Hamilton, James Madison, William C. Houston, George Read, Richard Bassett, Edmund Randolph, John Langdon, and Nathaniel Gorham).

71. *See, e.g.*, Simeon E. Baldwin, *The New Haven Convention of 1778*, *in* THREE HISTORICAL PAPERS READ BEFORE THE NEW HAVEN HISTORICAL SOCIETY 3, 37–38 (1882) (listing and discussing those interstate conventions commissioned to deal with issues of public credit).

72. CAPLAN, *supra* note 1, at 95–96 (citing Emer Vattel's then-popular work on international law).

73. *Id.*; *see also* THE FEDERALIST NO. 40, *supra* note 1, at 199 ("The powers of the convention ought, in strictness, to be determined, by an inspection of the commissions given to the members by their respective constituents.").

74. *E.g.*, 2 HOADLY, *supra* note 1, at 574 (reproducing Rhode Island's instructions to its delegates at the 1780 Philadelphia Convention, which dealt with price inflation).

75. *Id.*

76. *E.g.*, *id.* at 578 (reproducing a resolution of the 1780 Philadelphia convention, referring to it as a "meeting of the several states"). After the Constitution was ratified, early state applications applied similar nomenclature to a convention for proposing amendments. *See infra* note 78 and accompanying text.

77. *E.g.*, 17 JCC, *supra* note 1, at 790 (1910) (Aug. 29, 1780) (referring to the 1780 Boston Convention as a "convention of committees").

state applications for an amendments convention, and other legal and non-legal documents.

There is also a mass of material illuminating how the process was understood in years subsequent to the Founding. Although a convention for proposing amendments has never been held, state legislatures throughout the nineteenth and twentieth centuries issued hundreds of applications,[78] often amid intense public discussion. Also, courts frequently have ruled on Article V questions in ways that clarify the state-application-and-convention process.[79]

The remainder of this paper relies both on Founding and post-Founding evidence to deduce and explain the rules governing that procedure.

VIII. THE NATURE OF APPLICATIONS AND THE RULES GOVERNING THEM

A. The Nature of an Application

Article V provides that Congress shall call a convention for proposing amendments "on the Application of the Legislatures of two thirds of the several States."[80] Alexander Donaldson's *Universal Dictionary of the English Language,* published in 1763, contained the following relevant definitions of "application": "the act of applying one thing to another. The thing applied. The act of applying to any person, as a solicitor, or petitioner. . . . The address, suit, or request of a person. . . ."[81]

Other eighteenth-century definitions were not greatly different.[82] Nathaniel Bailey's dictionary defined the word as "the art of applying or addressing a person; also care, diligence, attention of the mind."[83] The same source defined "to apply" as "to put, set, or lay one thing to another, to have recourse to a thing or person, to betake, to give one's self up to."[84]

---

78. *See Convention Applications*, THE ARTICLE V LIBRARY: A PUBLIC RESOURCE FOR ARTICLE V RESEARCH, http://www.article5library.org/ (last visited May 5, 2011) (collecting hundreds of applications and related documents). Many applications are also collected at IMAGES OF ARTICLE V APPLICATIONS, http://www.article-5.org/file.php/1/Amendments/ (last visited May 5, 2011), although some of the documents labeled applications are documents of other kinds.

79. *See infra* text accompanying notes 305–323.

80. U.S. CONST. art. V.

81. ALEXANDER DONALDSON, AN UNIVERSAL DICTIONARY OF THE ENGLISH LANGUAGE (Edinburgh, 1763) (unpaginated) (defining "application").

82. *E.g.*, SAMUEL JOHNSON, 1 A DICTIONARY OF THE ENGLISH LANGUAGE (London, 8th ed. 1786) (unpaginated); THOMAS SHERIDAN, *supra* note 65.

83. NATHANIEL BAILEY, A UNIVERSAL ETYMOLOGICAL ENGLISH DICTIONARY (Edinburgh, Neill & Co., 25th ed. 1783) (unpaginated).

84. *Id.*

Thus, a state legislature's "Application" to Congress is the legislature's address to Congress requesting a convention.[85] Applications are adopted by legislative resolution.[86]

B. The Application Process is Not Subject to Normal Legislative Limitations, Such as Presentment to the Governor

Today, most governors must sign, and may veto, bills and many legislative resolutions. This gives them a share in the legislative power. Article V provides that applications are to be made by "the Legislatures of two thirds of the several States."[87] This raises the question of whether a state legislature operating under Article V includes the governor in states requiring the governor's signature on laws.[88] The evidence suggests that the answer is "no." Governors need not sign applications and may not veto them.[89]

The Constitution sometimes uses the term "legislature" to refer to the entire legislative process,[90] but on other occasions uses the term to designate the legislative assembly only. For example, the Guarantee Clause distinguishes "Application[s]" originating from "the Legislature" from those originating from "the Executive."[91] Similarly, election of United States Senators was entrusted to state legislatures without gubernatorial participation.[92]

Author Russell Caplan writes that the bitter colonial experience with royal governors argues that "legislature" in Article V refers to the representative assembly only.[93] His argument is strengthened by the 1789

---

85. Natelson, *Amending*, *supra* note 1, at 1.

86. *See generally* the applications at *Convention Applications*, *supra* note 78.

87. U.S. CONST. art. V.

88. Natelson, *Amending*, *supra* note 1, at 10.

89. CAPLAN, *supra* note 1, at 104–05; Natelson, *Amending*, *supra* note 1, at 10–11.

90. *E.g.*, U.S. CONST. art. I, § 4, cl. 1 ("The Times, Places and Manner of holding Elections for Senators and Representatives, shall be prescribed in each State by the Legislature thereof; but the Congress may at any time by Law make or alter such Regulations, except as to the Places of choosing Senators."); Smiley v. Holm, 285 U.S. 355, 372–73 (1932) (holding that this clause refers to the entire legislative process, including the governor); Davis v. Hildebrant, 241 U.S. 565, 568 (1916) (holding that this clause refers to the entire legislative process, including voter referendum).

91. U.S. CONST. art. IV, §4 ("The United States shall guarantee to every State in this Union a Republican Form of Government, and shall protect each of them against Invasion; and on Application of the Legislature, or of the Executive (when the Legislature cannot be convened) against domestic Violence.").

92. U.S. CONST. art. I, § 3, cl. 1 (assigning election of Senators to state legislatures); *cf.* U.S. CONST. art. I, § 3, cl. 2 (dividing between legislature and executive the responsibility for filling vacancies in the Senate).

93. CAPLAN, *supra* note 1, at 104.

amendment applications from New York and Virginia, both of which lacked the governor's signature.[94]

One might respond that because neither the governor of New York nor the governor of Virginia enjoyed a veto in 1789, they had no share in the legislative power—and that this explains why they did not sign their states' applications. However, the New York Constitution did vest a qualified veto, subject to a two thirds override, in a "council of revision" that included the governor.[95] Yet the council's approval does not appear on the application.[96] The Framers knew, moreover, that in Massachusetts the governor enjoyed a qualified veto,[97] and in soon-to-be-admitted Vermont, the governor's council held a suspensive veto.[98] Because the Constitution makes no mention of such powers, we can infer that the Framers' decision to mention only representative assemblies was deliberate.

In 1798, the Supreme Court held that Congress acts without the President when proposing amendments,[99] thereby implying that the same rule prevails at the state level. Newer case law likewise holds that Article V confers powers on named assemblies, not on the lawmaking apparatus per se.[100] In other words, resolutions pursuant to Article V, including those approving applications, are not considered legislative in nature.[101]

For the same reason, state constitutional provisions governing the *legislative* process do not apply to Article V applications. The courts have invalidated state constitutional rules mandating legislative super-majorities[102] and binding referenda[103] when such rules would apply to Article V resolutions. Restrictions on an Article V assembly's procedure

---

94. *Id.* at 104–05; H.R. JOURNAL, 1st Cong., 1st Sess. 29–30 (1789), *available at Convention Applications*, *supra* note 78.

95. N.Y. CONST. of 1777, art. III.

96. H.R. JOURNAL, 1st Cong., 1st Sess. 29–30 (1789), *available at Convention Applications*, *supra* note 78.

97. MASS. CONST. of 1780, ch. I, § I, art. II.

98. VT. CONST. of 1786, ch. II, § XVI.

99. Hollingsworth v. Virginia, 3 U.S. (3 Dall.) 378 (1798).

100. United States v. Sprague, 282 U.S. 716 (1931) (bestowing power on Congress); Hawke v. Smith, 253 U.S. 221 (1920) (bestowing power on state legislature).

101. *See supra* notes 99–100.

102. Dyer v. Blair, 390 F. Supp. 1291 (N.D. Ill. 1975) (Stevens, J.) (applying state constitutional requirement of a supermajority vote only because the legislature had freely adopted it when acting under Article V).

103. *See, e.g.*, Leser v. Garnett, 258 U.S. 130 (1922); *Hawke*, 253 U.S. 221; *see also* Prior v. Norland, 188 P. 729 (Colo. 1920); *In re* Opinion of the Justices, 167 A. 176 (Me. 1933); State *ex rel.* Tate v. Sevier, 62 S.W.2d 895 (Mo. 1933), *cert. denied*, 290 U.S. 679 (1933); *In re* Opinion of the Justices, 172 S.E. 474 (N.C. 1933); State *ex rel.* Donnelly v. Myers, 186 N.E. 918 (Ohio 1933). *But cf.* Kimble v. Swackhamer, 439 U.S. 1385, 1388, *appeal dismissed*, 439 U.S. 1041 (1978) (Rehnquist, J.) (permitting non-binding referendum).

are valid only if freely adopted by that assembly itself.[104] Correspondingly, an assembly is free to adopt its own procedures when discharging an Article V function.[105]

C. States May Rescind Applications

Some have argued that states cannot rescind applications, and that once adopted an application continues in effect forever, unless a convention is called.[106] This position is contrary to the principles of agency the Founders incorporated into the process.[107] An application is a deputation from the state legislature to Congress to call a convention.[108] Just as one may withdraw authority from an agent before the interest of a third party vests, so may the state legislature withdraw authority from Congress before the two thirds threshold is reached.[109]

Caplan demonstrates that the power of a state to rescind its resolutions, offers, and ratifications was well established by the time Article V was adopted, ending only when the culmination of a joint process was reached.[110] Just as a state may rescind ratification of a constitutional amendment any time before three fourths of the states have ratified,[111] it may also withdraw its application any time before two thirds of states have applied. At least one modern court has agreed.[112]

D. Applications Do Not Grow "Stale" with the Passage of Time

Some have argued that applications automatically become "stale" after an unspecified period of time, and no longer count toward a two thirds majority.[113] This argument is supported by a 1921 Supreme Court case, *Dillon v. Gloss*, suggesting that ratifications, to be valid, must be issued within a reasonable time of each other.[114]

As far as I have discovered, there is no evidence from the Founding Era or from early American practice implying that applications become stale automatically, or that Congress can declare them so. On the contrary,

---

104. *Dyer*, 390 F. Supp. at 1308.

105. *E.g.*, *id.* at 1307.

106. *See* Rees, *Amendment Process*, *supra* note 1, at 72 (discussing this position, but disagreeing).

107. *See* Natelson, *Amending*, *supra* note 1, at 15.

108. *Id.* at 19.

109. *See id.* at 73 (analogizing, as the Founders would have, to the law of nations).

110. CAPLAN, *supra* note 1, at 108–10.

111. Grover Rees, III, Comment, *Rescinding Ratification of Proposed Constitutional Amendments: A Question for the Court*, 37 LA. L. REV. 896, 896 (1977).

112. Idaho v. Freeman, 529 F. Supp. 1107 (D. Idaho 1981), *vacated as moot*, Carmen v. Idaho, 459 U.S. 809 (1982).

113. *See* Rees, *Amendment Process*, *supra* note 1, at 89.

114. Dillon v. Gloss, 256 U.S. 368, 375 (1921).

during the constitutional debates, participants frequently noted with approval the Constitution's general lack of time requirements in the amendment process.[115] Moreover, the ministerial nature of the congressional duty to call a convention[116] and Congress's role as the agent for those legislatures in this process,[117] suggests the opposite. Time limits are for principals, not agents, to impose. Therefore, if a state legislature believes its application to be stale, that legislature may rescind it.[118]

Events subsequent to *Dillon* support this inference. For example, the Supreme Court essentially has disavowed much of the "staleness" language in that case.[119] The universally-recognized adoption of the Twenty-Seventh Amendment, based on ratifications stretching over two centuries, points in the same direction.[120]

Even if *ratifications* become stale, it does not follow that *applications* do. The "staleness" discussion in *Dillon* was based partly on presumed congressional power to set ratification time limits as an incident of its power to choose one of two "Mode[s] of Ratification."[121] However, congressional authority to call a convention for proposing amendments is

---

115. *See Responses to An Old Whig I*, MASS. CENTINEL, Oct. 31, 1787, *reprinted in* 4 DOCUMENTARY HISTORY, *supra* note 1, at 179, 182 (1997):

> There is another argument I had nearly forgotten, and that is the degree of liberty admitted as to this power of revision in the new Constitution, which we have not expressed, even in that of Massachusetts—For the citizens of this Commonwealth are only permitted at a given time to revise their Constitution and then only if two thirds are agreed; but in the other case, the citizens of the United States can do it, without any limitation of time.

Id. For another writing celebrating the lack of time limits, see *Uncus*, MD. J., Nov. 9, 1787, *reprinted in* 14 DOCUMENTARY HISTORY, *supra* note 1, at 76 (1983) ("Should it be thought best at any time hereafter to amend the plan; sufficient provision for it is made in Art. 5, Sect. 3 . . . ." *Id.* at 81).

116. *See infra* note 257 and accompanying text.

117. *See* Natelson, *Amending*, *supra* note 1, at 15.

118. *See* Rees, *Amendment Process*, *supra* note 1, at 88 (arguing that the purpose of the process is such that each state legislature ought to control its own application); *cf.* CAPLAN, *supra* note 1, at 108–10 (explaining that the Founding-Era record suggests states have power to rescind their applications).

119. *See* Coleman v. Miller, 307 U.S. 433, 452–53 (1939) ("[I]t does not follow that, whenever Congress has not exercised that power [to fix a reasonable time for ratification], the Court should take upon itself the responsibility of deciding what constitutes a reasonable time and determine accordingly the validity of ratifications.").

120. *See* Paulsen, *supra* note 1, at 680 (citing the Justice Department's belief that because there was a "formal proposal by a two-thirds majority of both houses of Congress and [] formal ratifications of thirty-eight state legislatures[,]" time considerations were irrelevant).

121. Dillon v. Gloss, 256 U.S. 368, 376 (1921).

narrower than its authority over ratification: The latter is partly discretionary.[122] The former is purely ministerial.[123]

The Constitution prescribes no time period by which an application becomes "stale."[124] Hence, a decision as to whether a particular application is or is not "stale" is purely a matter of judgment.[125] As the Supreme Court has noted, the courts cannot make this judgment because they have no legal criteria by which to judge.[126] Leaving the decision to Congress would be the worst possible solution,[127] because doing so could defeat the central purpose of the state-application-and-convention process—to allow the states to bypass Congress. History strongly suggests that Congress would manipulate the period to interfere with the process. During the 1960s, for example, senators opposed to proffered amendments argued that all applications should be deemed stale (and therefore invalid) after a period of no more than two or three years![128] Because of the biennial schedule of many state legislatures, this would have effectively excised the state-application-and-convention process from the Constitution. Yet during the 1970s, when states balked at approving an amendment *Congress* had proposed, Congress purported to extend the ratification period from seven to ten years.[129]

In the final analysis, the only proper judge of whether an application is fresh or stale is the legislature that adopted it. Any time a legislature deems an application (or a ratification) outdated, the legislature may rescind it, as many have done.

---

122. *See* United States v. Sprague, 282 U.S. 716, 732–33 (1931) (discussing congressional discretion as to the mode of ratification).

123. *See infra* Part X.A–B. (discussing ministerial nature of call after applications).

124. CAPLAN, *supra* note 1, at 110.

125. *See id.* at 111 (arguing that "[i]n theory an application could remain effective . . . indefinitely.")

126. Coleman v. Miller, 307 U.S. 438, 453–54 (1939).

127. *See* Rees, *Amendment Process*, *supra* note 1, at 85 (discussing the conflict of interest in allowing Congress to determine time limits for ratification of amendments); *cf.* Paulsen, *supra* note 1, at 717 ("[T]he least defensible position would seem to be one of plenary congressional power . . . .").

128. CAPLAN, *supra* note 1, at 75–76 (quoting Senator Robert Kennedy).

129. *See* Idaho v. Freeman, 529 F. Supp. 1107 (D. Idaho 1981), *judgment vacated as moot by* Carmen v. Idaho, 459 U.S. 809 (1982) (concluding that "the congressional act of extending the time period for ratification [of the Equal Rights Amendment] was an improper exercise of Congress' authority under article V."); *see also* Grover Rees, III, *Throwing Away the Key: The Unconstitutionality of the Equal Rights Amendment Extension*, 58 TEX. L. REV. 875 (1980) (arguing that only the state legislatures have the power to extend their own ratifications).

IX. DEFINING THE SCOPE OF THE CONVENTION

A. Founding-Era Convention Practice Before the 1787 Convention

Perhaps no Article V question has been debated so fiercely, on so little evidence, as whether applying states may limit the scope of a convention for proposing amendments. A more complete view of the evidence tells us the answer is almost certainly "yes."

It is uncontroverted that state legislative applications may request a convention *unlimited* as to subject[130]—the sort of assembly the Founders, in imitation of international practice, called a plenipotentiary convention.[131] Many, however, have contended that the applying states do not have the complementary power of limiting the scope.[132] People so arguing deem an amendments convention a "constitutional convention,"[133] an inherently plenipotentiary body, enjoying power to propose any changes it wishes.[134] Others have asserted that it might be more than a proposing body: It could constitute itself a junta that could repeal the Bill of Rights, restore slavery, or otherwise radically alter our system of government.[135] How the convention could do these things without control of the military is never made clear.

The claim that any interstate convention is invariably a plenipotentiary "constitutional convention"—and therefore a potential "runaway"—first arose in the nineteenth century.[136] It has no Founding-Era pedigree and no basis in Founding-Era practice.

During that period, many conventions were held within individual colonies and states.[137] These included plenipotentiary gatherings that wrote state constitutions and otherwise erected new governments.[138] But they also

---

130. Such applications were submitted by New York in 1789, by Georgia in 1832, and by several other states in the run-up to the Civil War. Natelson, *First Century*, *supra* note 1, at 6, 8–13.

131. *See* CAPLAN, *supra* note 1, at 23. On the use of plenipotentiary conventions, see also *id.* at xx–xxi, discussing the scope of such conventions, and *id.* at 20, citing Hamilton's desire for calling a plenipotentiary convention to overhaul the Articles.

132. Ervin, *supra* note 1, at 881.

133. I have made that error in oral discussions of the Constitution; however, I have been in very good company. *See, e.g.*, Ervin, *supra* note 1, *passim*; Paulsen, *supra* note 1, at 738.

134. For an example of this approach, see Ralph M. Carson, *Disadvantages of a Federal Constitutional Convention*, 66 MICH. L. REV. 921, 922–24 (1968), arguing that once convened, attempts by Congress to impose limitations on subject matter would be of no avail.

135. CAPLAN, *supra* note 1, at vii–viii (quoting various public figures), 146–47 (quoting Theodore Sorensen).

136. *See id.* at xi–xv, 44, 47, 56, 60.

137. Natelson, *First Century*, *supra* note 1, at 3.

138. *Id.*

included conventions called for narrower purposes, such as state conventions for proposing amendments.[139] The Pennsylvania Constitution of 1776 and the Vermont Constitution of 1786, for example, both provided for limited amendments conventions, each restricted in its scope by a "council of censors."[140] The Massachusetts Constitution of 1780 provided for amendment by convention,[141] as did the Georgia Constitution of 1777. The latter instrument authorized the convention only to draft constitutional amendments whose gist had been prescribed by a majority of counties.[142]

---

139. *Id.*

140. PA. CONST. of 1776, § 47:

> The said council of censors shall also have power to call a convention, to meet within too [sic] years after their sitting, if there appear to them an absolute necessity of amending any article of the constitution which may be defective, explaining such as may be thought not clearly expressed, and of adding such as are necessary for the preservation of the rights and happiness of the people: But the articles to be amended, and the amendments proposed, and such articles as are proposed to be added or abolished, shall be promulgated at least six months before the day appointed for the election of such convention, for the previous consideration of the people, that they may have an opportunity of instructing their delegates on the subject.

Id.; *see also* VT. CONST. of 1786, ch. II, § XL (containing similar language).

141. MASS. CONST. of 1780, pt. II, ch. VI, art. X:

> In order the more effectually to adhere to the principles of the constitution, and to correct those violations which by any means may be made therein, as well as to form such alterations as from experience shall be found necessary, the general court which shall be in the year of our Lord [1795] shall issue precepts to the selectmen of the several towns, and to the assessors of the unincorporated plantations, directing them to convene the qualified voters of their respective towns and plantations, for the purpose of collecting their sentiments on the necessity or expediency of revising the constitution in order to amendments.
>
> And if it shall appear, by the returns made, that two-thirds of the qualified voters throughout the State, who shall assemble and vote in consequence of the said precepts, are in favor of such revision or amendment, the general court shall issue precepts, or direct them to be issued from the secretary's office, to the several towns to elect delegates to meet in convention for the purpose aforesaid.

Id.

142. GA. CONST. of 1777, art. LXIII:

> No alteration shall be made in this constitution without petitions from a majority of the counties . . . at which time the assembly shall order a convention to be called for that purpose, specifying the alterations to be made, according to the petitions preferred to the assembly by the majority of

The Georgia procedure may well have inspired the state-application-and-convention process of Article V.[143]

Some conventions were not limited to individual colonies or states, but were inter-colonial, interstate, or "federal."[144] The opening assembly of this sort in the Founding Era was the First Continental Congress (1774).[145] Its charge was plenipotentiary: "to consult and advise [i.e., deliberate][146] with the Commissioners or Committees of the several English Colonies in America, on proper measures for advancing the best good of the Colonies."[147] Between the First Continental Congress and the 1787 constitutional convention, there were at least ten other interstate gatherings.[148] All were limited to issuing recommendations, and none was plenipotentiary.[149] The broadest was probably the Springfield Convention of 1777, which was entrusted with issues of currency, monopoly and economic oppression, and interstate trade restrictions.[150] It was, however, limited formally to matters outside the authority of Congress.[151] Nearly as broad was the charge to the three-state Boston Convention of 1780, which was held to consider all aspects of the ongoing war.[152] The convention interpreted this charge liberally to include recommendations on trade and currency.[153]

The first Providence Convention (1776–77) was restricted to currency and defense measures.[154] Shortly thereafter, Congress recommended interstate conventions in York, Pennsylvania and Charleston, South Carolina, to consider the single subject of price-stabilization.[155] Because the Providence meeting had included the four New England states,[156] Congress recommended that New York, New Jersey, Pennsylvania, Maryland, Delaware, and Virginia meet at York and the Carolinas and Georgia

---

the counties as aforesaid.

Id.

143. Article XIX in the Committee of Detail's draft at the 1787 convention looked rather like the Georgia provision. *See* 2 FARRAND'S RECORDS, *supra* note 1, at 188.

144. *See* Natelson, *First Century*, *supra* note 1, at 3.

145. *See* Natelson, *Amending*, *supra* note 1, at 24 n.42.

146. On the meaning of "advise" as meaning in this context, to "deliberate," see NATELSON, ORIGINAL CONSTITUTION, *supra* note 1, at 70–72.

147. 1 JCC, *supra* note 1, at 18 (1904) (commission of Connecticut delegates).

148. Natelson, *Amending*, *supra* note 1, at 6; *see also* CAPLAN, *supra* note 1, at 16–26.

149. Natelson, *Amending*, *supra* note 1, at 6.

150. *Id.* at 24 n.44; *see also* CAPLAN, *supra* note 1, at 17–18.

151. 1 HOADLY, *supra* note 1, at 599.

152. *See* 3 HOADLY, *supra* note 1, at 559–64.

153. *See id.*

154. 1 HOADLY, *supra* note 1, at 585–86.

155. CAPLAN, *supra* note 1, at 17; 7 JCC, *supra* note 1, at 124–25 (1907) (Feb. 15, 1777) (reproducing the congressional calls).

156. Maine was then part of Massachusetts, and Vermont had not yet been admitted.

convene at Charleston.[157] It is unclear whether the Charleston meeting ever took place.[158] The York convention did meet; however, it did not issue a recommendation because of a tie vote among the states present.[159]

Interstate meetings at New Haven (1778) and Philadelphia (1780) also dealt only with price regulation.[160] The first Hartford Convention (1779) was empowered to address currency and trade,[161] and the second (1780) met "for the purpose of advising and consulting upon measures for furnishing the necessary supplies of men and provision for the army."[162] The second Providence Convention (1781) was entrusted only with recommending how to provide supplies to the army for a single year.[163]

The last of the limited-subject interstate gatherings is the most famous today. The Annapolis Convention of 1786 was to focus on "the trade and Commerce of the United States."[164] Its limited scope induced James Madison explicitly to distinguish it from a plenipotentiary convention.[165]

In sum, after the plenipotentiary First Continental Congress, all the interstate conventions were called to recommend solutions to one or more discrete, previously identified problems.[166] Today we probably would call them "task forces." For the most part, all remained within the scope of their calls.[167] If there was an exception, it was the assembly at Annapolis—and that exception was solely to express the "wish" and "opinion" that another convention be held to consider defects in the political system.[168] So, by

---

157. 7 JCC, *supra* note 1, at 124–25 (1907).

158. *See* CAPLAN, *supra* note 1, at 17 (asserting that "the Charleston convention never materialized.")

159. Byron W. Holt, *Continental Currency*, 5 SOUND CURRENCY, Apr. 1, 1898, at 81, 106–07 (discussing the York convention and other "price conventions"). *But see* 3 RICHARD HILDRETH, THE HISTORY OF THE UNITED STATES OF AMERICA 182 (1880) (claiming that the York convention did arrive at a price-fixing agreement).

160. 1 HOADLY, *supra* note 1, at 607 (New Haven); *Id.* at 572 (Philadelphia).

161. 2 HOADLY, *supra* note 1, at 562–63.

162. 3 HOADLY, *supra* note 1, at 565 (commission of New Hampshire delegates).

163. *Id.* at 575–76.

164. Proceedings of Commissioners to Remedy Defects of the Federal Government, *in* 1 THE DEBATES IN THE SEVERAL STATE CONVENTIONS, ON THE ADOPTION OF THE FEDERAL CONSTITUTION 116, 117 (2d ed. 1861) (Annapolis, Sept. 11, 1786) [hereinafter Proceedings of Commissioners], *available at* http://avalon.law.yale.edu/18th_century/annapoli.asp. Because only five states were present, the delegates voted not to proceed with their charge and suggested to Congress that it call a convention with a broader charge. *Id.* at 118; *cf.* Harmon, *supra* note 1, at 398 (pointing out that the Annapolis Convention was limited in nature).

165. *See* CAPLAN, *supra* note 1, at 23; see also *id.* at xx–xxi (explaining usage), 20 (quoting Hamilton).

166. *See id.* at 16–26.

167. The recommendation of a day of prayer by the first Providence Convention, 1 HOADLY, *supra* note 1, at 598–99, would have been seen by the founding generation as within the call.

168. *See* Proceedings of Commissioners, *supra* note 164, at 117–18.

1787, there had been ten interstate conventions, and not a single one had been a "runaway."[169]

B. Was the 1787 Federal Convention a "Runaway?"

Ann Diamond argues that reading Article V "so that it contemplates a constitutional convention that writes—not amends—a constitution, is often a rhetorical ploy to terrify sensible people."[170] For many years, central to that "ploy" has been the claim that the history of the 1787 federal convention (sometimes asserted to be the only federal convention ever held) illustrates how such an assembly can "run away." Directed by Congress to convene "for the sole and express purpose of revising the Articles of Confederation,"[171] the delegates (it is said) exceeded the limit Congress had placed on their authority. Instead, they scrapped the Articles and wrote an entirely new Constitution.[172]

It is true, of course, that they did write an entirely new Constitution; however, further examination reveals that the rest of this story is essentially false.

On September 14, 1786, the delegates to the Annapolis Convention recommended to the five states that had sent them—not to Congress—that those states coordinate with the other eight to call an assembly with authority to recommend changes to "render the *constitution* of the Federal Government adequate to the exigencies of the Union."[173] This resolution was merely a recommendation outside that assembly's powers, and as such, had no legal force.[174]

According to usages of the time, the term "constitution" usually did not denote a particular document, such as the Articles, but rather a governmental structure as a whole.[175] Particular documents traditionally had not been called "constitutions," but "instruments of government," "frames of government," or "forms of government." This explains why several of the early state constitutions described themselves in multiple terms.[176] In

---

169. *See* Natelson, *Amending*, *supra* note 1, at 6.

170. Diamond, *supra* note 1, at 137.

171. Report of Congressional Proceedings (Feb. 21, 1787), *in* 1 THE DEBATES IN THE SEVERAL STATE CONVENTIONS, ON THE ADOPTION OF THE FEDERAL CONSTITUTION 116, 117 (Philadelphia, J.B. Lippincott & Co. 2d ed. 1861).

172. *See, e.g.*, Douglas G. Voegler, *Amending the Constitution by the Article V Convention Method*, 55 N.D. L. REV. 355, 393 (1979).

173. Proceedings of Commissioners, *supra* note 164, at 118 (emphasis added).

174. *See supra* Part V.A.

175. For example, the 1786 edition of Johnson's dictionary contained only these political meanings of *constitution*: "Established form of government; system of laws and customs" and "Particular law; . . . establishment; institution." JOHNSON, *supra* note 82. The political definitions of *constitution* in the 1789 edition of Thomas Sheridan's dictionary were almost identical. SHERIDAN, *supra* note 65 (defining "constitution").

176. *See, e.g.*, DEL. CONST. of 1776, pmbl. ("Constitution, or System of Government");

other words, the Annapolis convention was suggesting changes necessary to render the federal *political system* "adequate to the exigencies" of the union.[177] However, the convention did suggest that any changes be approved by Congress and "afterwards confirmed by the Legislatures of every State."[178]

In the ensuing months, seven states provided for the appointment of delegates to a new convention in terms at least as broad as the Annapolis recommendation and without the proviso that any changes be approved by Congress and by every state.[179] On February 21, 1787, a committee of Congress recommended that Congress add its moral support to the idea.[180] This triggered the objection of the New York delegation, which offered substitute language limiting the recommendation only to amending the Articles.[181] Although Congress defeated the New York motion, it approved a compromise resolution offered by Massachusetts. This resolution also would have limited the scope of the Philadelphia convention:

> *Resolved* that in the opinion of Congress it is expedient that on the second Monday in May next a Convention of delegates who shall have been appointed by the several States be held at Philadelphia for the sole and express purpose of revising the Articles of Confederation and reporting to Congress and the several legislatures such alterations and provisions therein as shall when agreed to in Congress and confirmed by the States render the federal Constitution adequate to the exigencies of Government and the preservation of the Union.[182]

The limited nature of this resolution, "the sole and express purpose of revising the Articles of Confederation," constitutes the usual evidence cited for the narrow authority of the convention.[183] However, it does not prove what it is presented to prove, for it was not actually a legal call: Under the Articles of Confederation, Congress had no power to issue such a call, and

---

MASS. CONST. of 1780, pmbl. ("declaration of rights and frame of government as the constitution"); MD. CONST. of 1776, pmbl. ("Constitution and Form of Government"); VA. CONST. of 1776, tit. ("Constitution or Form of Government").

177. *See* Proceedings of Commissioners, *supra* note 164, at 118.

178. *Id.*

179. 3 FARRAND'S RECORDS, *supra* note 1, at 559 (reproducing the Virginia authorization, dated Oct. 16, 1786); *id.* at 563 (reproducing the New Jersey commission, dated Nov. 3, 1786); *id.* at 565–66 (reproducing Pennsylvania enabling legislation adopted Dec. 30, 1786); *id.* at 568 (showing that North Carolina elected its delegates in Jan., 1787); *Id.* at 571–72 (showing the New Hampshire resolution passing on Jan. 17, 1787); *id.* at 574 (showing the Delaware authorization as passing on Feb. 3, 1787); *id.* at 576–77 (reproducing the Georgia ordinance, adopted Feb. 10, 1787).

180. 32 JCC, *supra* note 1, at 71–72 (1936).

181. *Id.* at 72.

182. *Id.* at 73–74.

183. *Id.*

certainly none to define its scope.[184] Indeed, the words of the congressional resolution reflect its purely precatory nature—"in the opinion of Congress."[185] In other words, the congressional resolution, like that of the Annapolis gathering, was purely a recommendation.[186] States could participate or not, and under such terms as they wished. If they did so, as a matter of law, *the states*, not Congress, fixed the scope of their delegates' authority.[187] Congress had no authority whatsoever to restrict the authority the states gave their delegates.[188]

Six more states remained to be heard from. Rhode Island elected not to participate.[189] South Carolina, Connecticut, and Maryland stuck to the broader formula adopted by the initial seven.[190] Only Massachusetts[191] and New York[192] adopted the narrower congressional approach. But in Philadelphia, they were outnumbered ten states to two.[193]

---

184. ARTICLES OF CONFEDERATION OF 1781.

185. 32 JCC, *supra* note 1, at 74 (1936).

186. *Id.*

187. *Accord* CAPLAN, *supra* note 1, at 97; *see also* THE FEDERALIST NO. 40, *supra* note 1, at 199.

188. CAPLAN, *supra* note 1, at 97.

189. 3 FARRAND'S RECORDS, *supra* note 1, at 557–59 (listing the delegates at the convention).

190. *Id.* at 581, 585, 586 (reproducing the South Carolina, Connecticut, and Maryland credentials).

191. *Id.* at 584 (reproducing the Massachusetts credentials).

192. *Id.* at 579–80 (reproducing the New York credentials).

193. The wording of each commission varied somewhat, with some phrases repeating themselves. The relevant wording of each of the ten states' commissions was as follows: Connecticut:

> for the purposes mentioned in the said Act of Congress that may be present and duly empowered to act in said Convention, *and to discuss* upon such Alterations and Provisions agreeable to the general principles of Republican Government as they shall think proper to render the federal Constitution adequate to the exigencies of Government and, the preservation of the Union.

Id. at 585 (emphasis added). Delaware: "deliberating on, and discussing, such Alterations and further Provisions as may be necessary to render the Fœderal Constitution adequate to the Exigencies of the Union . . . ." *Id.* at 574. Georgia: "devising and discussing all such Alterations and farther Provisions as may be necessary to render the *Federal Constitution* adequate to the exigencies of the Union . . . ." *Id.* at 577 (italics in original). Maryland: "considering such Alterations and further Provisions as may be necessary to render the Fœderal Constitution adequate to the Exigencies of the Union . . . ." *Id.* at 586. New Hampshire: "devising & discussing all such alterations & further provisions as to render the federal Constitution adequate to the Exigencies of the Union . . . ." *Id.* at 572. New Jersey: "taking into Consideration the state of the Union, as to trade and other important objects, and of devising such other Provisions as shall appear to be necessary to render the Constitution of the Federal Government adequate to the exigencies thereof." *Id.* at 563.

At the convention itself, the Massachusetts and New York delegates were in a quandary. Elbridge Gerry of Massachusetts questioned the convention's authority to recommend changes extending beyond the Articles,[194] and ultimately refused to sign. His colleague Caleb Strong was forced to return home to tend a sick wife, so he was spared from having to make a choice.[195] The other two Bay State delegates, Rufus King and Nathaniel Gorham, both participated and added their names.

Of the three New Yorkers, two left early.[196] The third New Yorker, Alexander Hamilton, was not of a particularly scrupulous cast, and he fitfully participated and finally signed the Constitution—although in fairness, it should be pointed out that Hamilton signed only as an individual; because of the departure of his colleagues he no longer was an official representative of his state.

In addition, the credentials of the five Delaware signers, while broad enough to authorize scrapping most of the Articles, did limit the delegates in one particular: they were not to agree to any changes that altered the rule that "in the United States in Congress Assembled each State shall have one Vote."[197] Because the new bicameral Federal Congress was a very different entity with a very different role than the Articles of Confederation's unicameral "United States in Congress Assembled,"[198] the Delaware delegates could argue that they had remained within the strict letter of their commission.[199] Even if they had not, at most only seven or eight of the

---

North Carolina: "for the purpose of revising the Fœderal Constitution . . . To hold, exercise and enjoy the appointment aforesaid, with all Powers, Authorities and Emoluments to the same belonging or in any wise appertaining . . . ." *Id.* at 567–68.
Pennsylvania:

> "to meet such Deputies as may be appointed and authorized by the other States, to assemble in the said Convention at the City aforesaid, and to join with them in devising, deliberating on, and discussing, all such alterations and further Provisions, as may be necessary to render the fœderal Constitution fully adequate to the exigencies of the Union . . . ."

Id. at 565–66. South Carolina: "devising and discussing all such Alterations, Clauses, Articles and Provisions, as may be thought necessary to render the Fœderal Constitution entirely adequate to the actual Situation and future good Government of the confederated States . . . ." *Id.* at 581. Virginia: "devising and discussing all such Alterations and farther Provisions as may be necessary to render the Fœderal Constitution adequate to the Exigencies of the Union . . . ." *Id.* at 560.

194. 1 FARRAND'S RECORDS, *supra* note 1, at 42–43.
195. 3 FARRAND'S RECORDS, *supra* note 1, at 590.
196. *Id.* at 588, 590.
197. *Id.* at 574–75.
198. U.S. ARTICLES OF CONFEDERATION of 1781, art. 5, para. 4.
199. 3 FARRAND'S RECORDS, *supra* note 1, at 574–75

thirty-nine signers exceeded their authority,[200] leaving one well short of the charge that the Philadelphia convention as a whole was a "runaway." The overwhelming majority of delegates to the 1787 convention, like the delegates to other Founding-Era interstate conventions, remained within the scope of their power.

In any event, the recommendation of the convention was only a recommendation: non-binding and utterly without independent legal force—a recommendation such as any agent was entitled to make.[201] The convention did not impose its handiwork on the states or on the American people. States could approve or reject as they liked, with no state bound that refused to ratify.[202] In fact, unlike a convention for proposing amendments, the Philadelphia assembly was not even entitled to have its decisions transmitted to the states or considered by them.[203] James Wilson summed up the delegates' position: "authorized to *conclude nothing*, but . . . at liberty to *propose any thing*."[204]

Thus, we can glean the following from the history of Founding-Era interstate conventions: Most were limited to specific subjects. All honored the scope of their commissions. Construed most unfavorably to the delegates, the history shows that some of them, when far from home without modern means of communicating with their superiors, chose to interpret their authority liberally and make non-binding recommendations rather than accomplish nothing. But this history offers no evidence to suggest that conventions for proposing amendments cannot be limited, and almost none to suggest they are likely "runaways."

C. Other Evidence that Applications Can Limit the Convention's Agenda

The prevalence of limited-purpose conventions during the Founding Era places the evidentiary burden on those who contend that an Article V convention is somehow illimitable. There is no way they can carry that burden, because almost all the Founding-Era evidence is against them.

The first kind of evidence is the purpose of the state-application-and-convention procedure: to serve as an effective congressional bypass.[205] Without the power to specify the kinds of amendments they wanted, the states could apply for a convention only if they wished to open the entire Constitution for reconsideration. There is a strong presumption against an

---

200. *Id.* at 574, 579–80, 584 (reproducing the Delaware, New York, and Massachusetts credentials).

201. *See supra* note 103; Part V.

202. U.S. CONST. art. VII.

203. 32 JCC, *supra* note 1, at 74 (1936).

204. 1 FARRAND'S RECORDS, *supra* note 1, at 253. Wilson's use of "propose" here means "recommend." This should not be confused with the technical term employed in Article V. *See* U.S. CONST. art. V.

205. U.S. CONST. art. V.

interpretation of a constitutional provision that would undercut the value of the provision, and impair its principal purpose.

The second kind of evidence is the treatment of conventions in the constitutional text. The text authorizes state conventions for ratifying the Constitution,[206] state conventions for ratifying amendments,[207] and federal conventions for proposing amendments.[208] Both of the first two were clearly limited in nature: No sane person would suggest that a state ratifying convention, for example, also has inherent authority unilaterally to re-write the state constitution. As for the convention for proposing amendments, the text placed certain topics outside the amendment process[209] and therefore outside its competence, thereby affirming its limited nature.[210]

The third kind of evidence consists of the records of the 1787 drafting convention. Although other writers seem to have overlooked this point,[211] the fact is that the Philadelphia delegates actively considered providing for amendment by plenipotentiary conventions, but rejected that approach. Edmund Randolph's initial sketch in the Committee of Detail[212] and the first draft of the eventual Constitution by that committee[213] both contemplated plenipotentiary conventions that would prepare and adopt amendments. During the proceedings, the delegates opted instead for an assembly that would merely propose.[214] Later on, Roger Sherman moved to revert to a plenipotentiary formula, but his motion was soundly rejected.[215]

---

206. U.S. CONST. art. VII.

207. U.S. CONST. art. V.

208. *Id.*

209. U.S. CONST. art. V (slave trade and apportionment of taxes before 1808; equal suffrage of states in the Senate).

210. 32 JCC, *supra* note 1, at 74 (1936).

211. *But see* Harmon, *supra* note 1, at 399.

212. 2 FARRAND'S RECORDS, *supra* note 1, at 148. According to Randolph's version, "5. (An alteration may be effected in the articles of union, on the application of two thirds *nine* <2/3d> of the state legislatures <by a Convn.>) <on appln. of 2/3ds of the State Legislatures to the Natl. Leg. they call a Convn. to revise or alter ye Articles of Union>." *Id.*

213. *Id.* at 188 ("On the application of the Legislatures of two thirds of the States in the Union, for an amendment of this Constitution, the Legislature of the United States shall call a Convention for that purpose.").

214. *Id.* at 558.

215. *Id.* at 630. The text explains that Mr. Sherman's motion was rejected:

> Mr Sherman moved to strike out of art. V. after "legislatures" the words "of three fourths" and so after the word "Conventions" leaving future Conventions to act in this matter, like the present Conventions according to circumstances.
>
> On this motion
>
> N— H— divd. Mas— ay— Ct ay. N— J. ay— Pa no. Del— no. Md no. Va no. N. C. no. S— C. no. Geo— no. [Ayes — 3; noes — 7; divided — 1.]

Id.

Principal credit for replacing the plenipotentiary approach with the convention for proposing amendments belongs to Elbridge Gerry.[216] He objected to a draft authorizing the convention to modify the Constitution without state approval.[217] The other delegates agreed, considering first a requirement that any amendments the convention adopted be approved by two thirds of the states, but later strengthening that requirement to three fourths.[218] During the process Madison wondered why, if states applied for one or more amendments, a convention was even necessary: He "did not see why Congress would not be as much bound to propose amendments applied for by two thirds of the States as to call a Convention on the like application."[219] In other words, Madison referred to the states "appl[ying]" for amendments, with either the convention or Congress being "bound to propose" them.[220] Nevertheless, the delegates preferred that a body separate

---

216. *Id.* at 557–58.

217. *Id.* Mr. Gerry questioned the wisdom of the draft's provision:

> Mr Gerry moved to reconsider art XIX. viz, "On the application of the Legislatures of two thirds of the States in the Union, for an amendment of this Constitution, the Legislature of the U. S. shall call a Convention for that purpose."
>
> This Constitution he said is to be paramount to the State Constitutions. It follows, hence, from this article that two thirds of the States may obtain a Convention, a majority of which can bind the Union to innovations that may subvert the State-Constitutions altogether. He asked whether this was a situation proper to be run into–

Id.

218. *Id.* at 558–59. The requirement was changed to three fourths:

> On the motion of Mr. Gerry to reconsider
>
> N. H. divd. Mas. ay— Ct. ay. N. J— no. Pa ay. Del. ay. Md. ay. Va. ay. N— C. ay. S. C. ay. Geo. ay. [Ayes — 9; noes — 1; divided — 1.]
>
> Mr. Sherman moved to add to the article "or the Legislature may propose amendments to the several States for their approbation, but no amendments shall be binding until consented to by the several States"
>
> Mr. Gerry 2ded. the motion
>
> Mr. Wilson moved to insert "two thirds of" before the words "several States" — on which amendment to the motion of Mr. Sherman
>
> N. H. ay. Mas. <no> Ct. no. N. J. <no> Pa. ay— Del— ay Md. ay. Va. ay. N. C. no. S. C. no. Geo. no. [Ayes — 5; noes — 6.]
>
> Mr. Wilson then moved to insert "three fourths of" before "the several Sts" which was agreed to nem: con:

Id.

219. 2 FARRAND'S RECORDS, *supra* note 1, at 629–30; *accord* Harmon, *supra* note 1, at 398–401 (discussing this remark in wider context).

220. 2 FARRAND'S RECORDS, *supra* note 1, at 629–30.

from Congress perform the drafting, and the final wording, penned primarily by Madison, reflected that sentiment.[221]

The fourth kind of evidence consists of comments from Federalists promoting the Constitution during the ratification debates. Among those were some emphasizing the essential equality of Congress and the states in proposing amendments. In *Federalist No.43*, for example, Madison wrote that the Constitution "equally enables the general and the State governments to originate the amendment of errors."[222] Similarly, "A Native of Virginia" wrote that "whenever two-thirds of both Houses of Congress, or two-thirds of the State Legislatures, shall concur in deeming amendments necessary, a general Convention shall be appointed, the result of which, when ratified by three-fourths of the Legislatures, shall become part of the Federal Government."[223] The "Native" erred in saying that congressional action would provoke a convention, but his core message was the same as Madison's: As far as amendments were concerned, Congress and the states were on equal ground.[224]

Technically, of course, Congress and the states were not, and are not, on completely equal ground as far as amendments are concerned. Congress may propose directly, while the states must operate through a convention.[225] Still, the Federalist representations of equality suggest that in construing Article V, preference should be given to interpretations that raise the states

---

221. 2 FARRAND'S RECORDS, *supra* note 1, at 559. Madison suggested the adopted wording:

> Mr. Madison moved to postpone the consideration of the amended proposition in order to take up the following,
>
> "The Legislature of the U— S— whenever two thirds of both Houses shall deem necessary, or on the application of two thirds of the Legislatures of the several States, shall propose amendments to this Constitution, which shall be valid to all intents and purposes as part thereof, when the same shall have been ratified by three fourths at least of the Legislatures of the several States, or by Conventions in three fourths thereof, as one or the other mode of ratification may be proposed by the Legislature of the U. S:"
>
> Mr. Hamilton 2ded. the motion.
>
> . . . .
>
> On the question On the proposition of Mr. Madison & Mr. Hamilton as amended
>
> N. H. divd. Mas. ay. Ct. ay. N. J. ay. Pa. ay. Del. no. Md. ay. Va ay. N. C. ay S. C. ay. Geo. ay. [Ayes — 9; noes — 1; divided — 1.]

Id.

222. THE FEDERALIST NO. 43, *supra* note 1, at 228.

223. *A Native of Virginia: Observations upon the Proposed Plan of Federal Government, 2 April*, VA. GAZETTE, Apr. 2, 1788, *reprinted in* 9 DOCUMENTARY HISTORY, *supra* note 1, at 655, 689 (1990).

224. *Id.* at 689.

225. U.S. CONST. art. V.

toward the congressional level and treat the convention as their joint assembly.[226] This, in turn, suggests that if Congress may specify a subject when it proposes amendments, the states may do so as well.

A fifth kind of evidence also comes from the ratification-era record. These reveal unambiguous understandings, both among Federalists and Anti-Federalists, that (1) the convention was not plenipotentiary but rather that (2) the applying states could—in fact, usually would—specify particular subject-matter at the beginning of the process. As Hamilton wrote in *Federalist No. 85*, "every amendment to the Constitution, if once established, would be a single proposition, and might be brought forward singly. . . . And consequently, whenever nine, or rather ten States, were united in the desire of a particular amendment, that amendment must infallibly take place."[227] Hamilton's reference to nine states represented the two thirds then necessary to force a convention, and his reference to ten states represented the three fourths necessary to ratify the convention's proposals.[228] Later in the same paper, he referred to "two thirds or three fourths of the State legislatures" uniting in particular amendments.[229]

Similarly, George Washington understood that applying states would specify the convention subject-matter.[230] In April, 1788, he wrote to John Armstrong that "a constitutional door is open for such amendments as shall be thought necessary by nine States."[231] When explaining that Congress could not block the state-application-and-convention procedure, the influential Federalist writer Tench Coxe did so in these words:

> If two thirds of those legislatures require it, Congress *must* call a general convention, even though they dislike the proposed amendments, and if three fourths of the state legislatures or conventions approve such proposed amendments, they become *an actual and binding part of the constitution,* without any possible interference of Congress.[232]

---

226. THE FEDERALIST NO. 43, *supra* note 1, at 228.

227. THE FEDERALIST NO. 85, *supra* note 1, at 456.

228. *Id.*

229. *Id.* at 457. At the Massachusetts ratifying convention, Charles Jarvis similarly spoke of "nine states" approving particular amendments, but Dr. Jarvis seems to have been operating on the assumption that Rhode Island would not ratify. 2 ELLIOT'S DEBATES, *supra* note 1, at 130 (also referring to a total of "twelve states"). In that event, application would have to be by eight states (of 12) and ratification by nine.

230. Letter from George Washington to John Armstrong (Apr. 25, 1788) (on file with the University of Virginia Library), *available at* http://etext.virginia.edu/etcbin/toccer-new2?id=WasFi29.xml&images=images/modeng&data=/texts/english/modeng/parsed&tag=public&part=359&division=div1.

231. *Id.*

232. *A Pennsylvanian to the New York Convention*, PA. GAZETTE, June 11, 1788, *reprinted in* 20 DOCUMENTARY HISTORY, *supra* note 1, at 1139, 1142–43 (2004).

The passage reveals an assumption that states would make application explicitly to promote particular amendments.

Madison, Hamilton, Washington, and Coxe were all Federalists, but on this issue their opponents agreed. An Anti-Federalist writer, "An Old Whig," argued that amendments were unlikely:

> [T]he legislatures of two thirds of the states, must agree in desiring a convention to be called. This will probably never happen; but if it should happen, then the convention may agree to the amendments or not as they think right; and after all, three fourths of the states must ratify the amendments. . . .[233]

("The amendments" here presumably means the amendments proposed in advance of the convention.) Another Anti-Federalist, Abraham Yates, Jr., wrote, "We now Cant get the Amendments unless 2/3 of the States first Agree to a Convention And as Many to Agree to the Amendments—And then 3/4 of the Several Legislatures to Confirm them[.]"[234]

The Ratifiers shared the understanding that an amendments convention would not be plenipotentiary and that the applying states generally would limit the subjects addressed.[235] The future Chief Justice John Marshall distinguished at the Virginia ratifying convention between the gathering at Philadelphia and the more narrow amending procedure: "The difficulty we find in amending the Confederation," he said, " will not be found in amending this Constitution. Any amendments, in the system before you, will not go to a radical change; a plain way is pointed out for the purpose."[236] This mirrored the view of Madison, shortly before he became a Virginia convention delegate. In a November, 1788 letter to George Lee Turberville, he had recognized differences between a convention that considers "first principles,"[237] which "cannot be called without the unanimous consent of the parties who are to be bound by it," and a Convention for proposing amendments, which could be convened under the "forms of the Constitution" by "previous application of 2/3 of the State legislatures."[238] At the North Carolina ratifying convention James Iredell, a

---

233. Letter from *An Old Whig I*, PHILA. INDEP. GAZETTEER (Oct. 12, 1787), *reprinted in* 13 DOCUMENTARY HISTORY, *supra* note 1, at 376–77 (1981).

234. Letter from Abraham Yates, Jr., to William Smith (Sept. 22, 1788), *reprinted in* 23 DOCUMENTARY HISTORY, *supra* note 1, at 2474 (2009).

235. *See* FARRAND'S RECORDS, *supra* note 1, at 476.

236. 3 ELLIOT'S DEBATES, *supra* note 1, at 234.

237. 2 FARRAND'S RECORDS, *supra* note 1, at 476 (reporting Madison as saying, "The people were in fact, the fountain of all power, and by resorting to them, all difficulties were got over. They could alter constitutions as they pleased. It was a principle in the Bills of rights that *first principles* might be resorted to."). That Madison was referring to an unlimited convention when he spoke of "first principles" is confirmed by his use of the phrase at the federal convention.

238. Letter from James Madison to George Lee Turberville (Nov. 2, 1788), *in* 11 THE

Federalist who, like Marshall, later sat on the United States Supreme Court, also emphasized the limited nature of an amendments convention by pointing out that its proposals had to be approved by three fourths of the states.[239]

Other statements by the Ratifiers show that they believed that the states, more often than not, would determine the subject matter to be considered in an amendments convention.[240] In Virginia, Anti-Federalists argued that before the Constitution was ratified a new plenary constitutional convention should be called to re-write the document and add a bill of rights.[241] A Federalist leader, George Nicholas, rejoined that it made more sense to ratify first, and then employ Article V's state-application-and-convention route:

> On the application of the legislatures of two thirds of the several states, a convention is to be called to propose amendments, which shall be a part of the Constitution when ratified by the legislatures of three fourths of the several states, or by conventions in three fourths thereof. It is natural to conclude that those states who will apply for calling the convention will concur in the ratification of the proposed amendments.[242]

Of course, such a conclusion would be "natural" only if the convention was expected to stick to the agenda of the states that "apply for calling the convention."[243] That there would be such an agenda was confirmed by what Nicholas said next, predicting a future plenary convention:

PAPERS OF JAMES MADISON 330–31 (Robert A. Rutland & Charles F. Hobson eds., 1977).

Professor Walter E. Dellinger has argued that letters written about the same time by Madison to Philip Mazzei and George Eve suggest that Madison thought the states could not limit the convention subject matter. Dellinger, *supra* note 1, at 1643 n.46. The letters actually say nothing about the issue; they merely express fear that delegates hostile to the Constitution might abuse a convention. Letter from James Madison to Phillip Mazzei (Dec. 10, 1788), *in* 11 THE PAPERS OF JAMES MADISON 388, 404 (Robert A. Rutland & Charles F. Hobson eds., 1977). Indeed, the portion Professor Dellinger quoted from the Mazzei letter cuts the other way: "The object of the anti-federalists is to bring about another General Convention, which would either agree on nothing as would be agreeable to some, and throw everything into confusion; or expunge from the Constitution parts which are held by its friends to be essential to it." *Id.* at 389. The reason this cuts the other way is that since several ratifying conventions had proposed amendments that would "expunge" from the Constitution parts "held by its friends to be essential to it," a convention proposing such changes would be following state instructions. *Id.*

239. 4 ELLIOT'S DEBATES, *supra* note 1, at 177 (quoting Iredell at the North Carolina ratifying convention).

240. 3 ELLIOT'S DEBATES, *supra* note 1, at 101–02.

241. *Id.*

242. *Id.*

243. *Id.* at 102.

> There are strong and cogent reasons operating on my mind, that the amendments, which shall be agreed to by those states, will be sooner ratified by the rest than any other that can be proposed. The [ratifying] conventions which shall be so called will have their deliberations confined to a few points; no local interest to divert their attention; nothing but the necessary alterations. They will have many advantages over the last [plenary] Convention. No experiments to devise; the general and fundamental regulations being already laid down.[244]

During the ratification era, there seems to have been little dissent to the understanding that the applying states would set the agenda.[245] The belief was so widespread it sometimes led to the assumption that the states rather than the convention would do the proposing. We have seen Tench Coxe suggest as much in the extract quoted above.[246] Another instance occurred at the Virginia ratifying convention, where Patrick Henry observed that, "Two thirds of the Congress, or of the state legislatures, are necessary even to propose amendments."[247] A Federalist writing under the name of Cassius asserted that "the states may propose any alterations which they see fit, and that Congress shall take measures for having them carried into effect."[248]

That the founding generation thought that way is demonstrated by the procedure they followed in adopting the Bill of Rights—a procedure very close to the one initially proposed by Edmund Randolph at the federal convention.[249] As a first step, seven states, although through their ratifying

---

244. *Id.* (emphasis added).

245. CAPLAN, *supra* note 1, at 139–40. Caplan reproduces three comments from the latter part of 1788, suggesting that it would be better for Congress to propose amendments than for a convention to do so, because the latter might run out of control. *Id.* Two were anonymous pieces in Maryland newspapers appearing within three days of each other, perhaps by the same author, designed to combat Anti-Federalist demands for a second convention. *Id.* However, the second convention the Anti-Federalists were advocating would have been plenipotentiary or, if held under Article V, unrestricted by subject matter. *Id.* at 140. The third item was a letter from Paris by Thomas Jefferson, referring specifically to New York's efforts, furthered by a circular letter from Governor George Clinton, also for an unrestricted convention. *Id.*

246. *See supra* note 232 and accompanying text.

247. 3 ELLIOT'S DEBATES, *supra* note 1, at 49; *see also* 3 FARRAND'S RECORDS, *supra* note 1, at 367–68 (reproducing memoranda by George Mason stating that "the constn as agreed at first was that amendments might be proposed either by Congr. or the [state] legislatures . . . ." After a change, "they then restored it as it stood originally.").

248. *Cassius VI*, MASS. GAZETTE, Dec. 25, 1787, *reprinted in* 5 DOCUMENTARY HISTORY, *supra* note 1, at 511–12 (1998).

249. 2 FARRAND'S RECORDS, *supra* note 1, at 479 ("Mr. Randolph stated his idea to be . . . that the State Conventions should be at liberty to propose amendments to be submitted to another General Convention which may reject or incorporate them, as shall be judged proper."). Later, Mr. Randolph restated his proposal, but this time with a second plenary convention having "full power to settle the Constitution finally." *Id.* at 561. He restated the proposal yet again later. *Id.* at 564, 631.

conventions rather than their legislatures, adopted sample amendments for consideration by a later proposing body.[250] Sam Adams urged this step to the Massachusetts ratifying convention, saying the states should "particularize the amendments necessary to be proposed."[251] Next, an Article V convention—or Congress, if it acted quickly enough, as it did—would choose among the state suggestions, draft the actual amendments, and send them to the states for ratification or rejection.[252] Finally, the states would either ratify or reject.

A sixth and final category of evidence on this subject consists of early practice—both practice early enough to shed light on the views of the Founders and practice that revealed a later understanding of the Founders' plan.[253] The first item comes from 1789, before all the states had ratified the Constitution. Early that year, Virginia and New York both presented applications to Congress. The New York application was clearly plenary, but the Virginia application asked that

> a convention be immediately called . . . with full power to take into their consideration *the defects of the Constitution that have been suggested by the State Conventions*, and report such amendments thereto as they shall find best suited to promote our common interests, and secure to ourselves and our latest posterity the great and unalienable rights of mankind.[254]

The language renders it likely that Virginia lawmakers intended the convention to select its proposals from among the topics suggested by the ratifying conventions.

The next applications arose out of the nullification crisis of the early 1830s. They were the 1832 applications from South Carolina and Georgia and the 1833 application from Alabama. Those of both South Carolina[255] and Alabama[256] called for a convention to address particular subjects. So

---

250. *See generally* 2 ELLIOT'S DEBATES, *supra* note 1 (outlining the occurrences at the seven state conventions).

251. *Id.* at 124.

252. Congress did propose one provision not on any of the states' lists: the Takings Clause—but of course Congress, unlike an Article V convention, had plenary power to propose amendments. The Takings Clause may have been an effort to respond to a ratification-era interpretation of the federal Ex Post Facto Clause that Madison believed was narrower than initially intended. NATELSON, ORIGINAL CONSTITUTION, *supra* note 1, at 157–58; *see also* Robert G. Natelson, *Statutory Retroactivity: The Founders' View*, 39 IDAHO L. REV. 489, 523 (2003).

253. *See* NATELSON, ORIGINAL CONSTITUTION, *supra* note 1, at 40 (explaining that evidence of the original meaning of the unamended Constitution is of limited value if arising later than May 29, 1790). Later evidence is usually merely evidence of *later* understandings.

254. Natelson, *Amending*, *supra* note 1, at 14 (emphasis added).

255. H. JOURNAL, 22d Cong., 2d Sess. 219–20 (1833). (reproducing the South Carolina application).

256. *Id.* at 361–62 (reproducing the Alabama application).

also did an 1864 application from Oregon, which was targeted at slavery.[257] Ensuing decades witnessed a veritable flood of single-subject applications on such topics as direct election of U.S. Senators and control of polygamy.[258]

Thus, the historical evidence pretty well disproves the view of a few writers that state applications specifying subject matter are void or that conventions for proposing amendments were to be governed by rules different from those applied to other Founding-Era conventions.[259] Case law on the subject is scanty, but what is available is consistent with the power of legislatures to limit the convention's subject.[260]

X. THE CONVENTION CALL AND SELECTION OF DELEGATES

A. Congress as a (Limited) Agent of the States

As noted above, key to understanding the intended operation of Article V—and the Constitution generally—is understanding how fiduciary principles were to govern that operation.[261]

Under the Confederation, Congress generally had been the fiduciary, specifically the agent, of the *states*. Under the Constitution, Congress became, for most purposes, the agent of the American *people*.[262] However, the congressional role in calling an amendments convention differs importantly from its usual role; in calling the convention and sending its proposals to the states, Congress acts as a ministerial agent of the *state legislatures*[263]—a conclusion buttressed by other evidence discussed later.[264] In this respect, the Framers retained the Confederation way of

---

257. Natelson, *First Century*, *supra* note 1, at 13.). It was thus erroneous to claim, as some writers have, that, "For a century following the Constitutional Convention in 1787, the only applications submitted by state legislatures under Article V contemplated conventions that would be free to determine their own agendas." Dellinger, *supra* note 1, at 1623 (citing Black, *Amending*, *supra* note 1, at 202). Black, however, does not fully support the statement. *See* Black, *Amending*, *supra* note 1, at 202.

258. *See* Natelson, *First Century*, *supra* note 1, at 8–14,19–21.

259. *E.g.*, Charles L. Black, *Amending*, *supra* note 1, at 198–99.

260. *E.g.*, *In re* Opinion of the Justices, 172 S.E. 474, 477 (N.C. 1933) (concluding that a state may limit authority of a ratifying convention); *see also* Opinion of the Justices to the Senate, 366 N.E.2d 1226, 1229 (Mass. 1977) (holding that a single-subject application is a valid application, and although refusing to hold that it would restrict the convention, noting that the Founders expected the states to specify subject-matter in their applications).

261. *See supra* Part V.A.

262. NATELSON, ORIGINAL CONSTITUTION, *supra* note 1, at 41–44.

263. *See* CAPLAN, *supra* note 1, at 94; *see also* Rees, *Amendment Process*, *supra* note 1, at 92 (referring to "Congress's ministerial duty to call a convention requested by the State legislatures").

264. *See infra* Part X.B.

doing things. They did so because of the need for an amendment procedure through which the states could bypass congressional discretion.

During the 1787 convention, the initial Virginia Plan called for an amendments convention to be triggered only by the states, leaving Congress without power to call one on its own motion.[265] The delegates altered this to allow only Congress to call an amendments convention.[266] George Mason then pointed out that if amendments were made necessary by Congress's own abuses, Congress might block them unless the Constitution contained an alternative route.[267] Accordingly, "Mr. Govr. Morris & Mr. Gerry moved to amend the article so as to require a Convention on application of 2/3 of the Sts."[268] If the proper number of states applied, Congress had no choice in the matter; it was constrained to do their bidding.[269]

As an agent for states in making the call, Congress was expected to follow rules of fiduciary law, including the duty to treat all of its principals (the state legislatures) impartially. It followed, for example, that Congress could not prescribe procedures that gave some states more power at the convention than others.

B. Congress's Role in Calling the Convention

Because the state-application-and-convention procedure was designed to bypass congressional discretion, the congressional discretion had to be strictly limited. In other words, it had to be chiefly clerical—or, to use the legal term, "ministerial."[270] On this point, Professor William W. Van Alstyne summarized his impressions of the history of Article V:

---

265. 2 FARRAND'S RECORDS, *supra* note 1, at 466–67.

266. *Id.* at 467–68. ("Art: XIX taken up. Mr. Govr. Morris suggested that the Legislature should be left at liberty to call a Convention, whenever they please. The art: was agreed to nem: con:").

267. 2 FARRAND'S RECORDS, *supra* note 1, at 629.

> Col: Mason thought the plan of amending the Constitution exceptionable & dangerous. As the proposing of amendments is in both the modes to depend, in the first immediately, and in the second, ultimately, on Congress, no amendments of the proper kind would ever be obtained by the people, if the Government should become oppressive, as he verily believed would be the case.

Id. at 629.

268. *Id.*

269. *See supra* notes 261–267 and accompanying text.

270. *See* Bruce M. Van Sickle & Lynn M. Boughey, *A Lawful and Peaceful Revolution, Article V and Congress' Present Duty to Call a Convention for Proposing Amendments*, 14 HAMLINE L. REV. 1, 41 (1990) (stating that Congress's role must, as much as possible, be merely mechanical or ministerial rather than discretionary); *see also* Rees, *Amendment Process*, *supra* note 1, at 92 (referring to the congressional call as "ministerial").

> The various stages of drafting through which article V passed convey an additional impression as well: that the state mode for getting amendments proposed was not to be contingent upon any significant cooperation or discretion in Congress. Except as to its option in choosing between two procedures for ratification, either "by the Legislatures of three fourths of the several States, or by Conventions in three fourths thereof," Congress was supposed to be mere clerk of the process convoking state-called conventions.[271]

As the writer of a Harvard Law Review Note observed, "any requirement imposed by Congress which is not necessary for Congress to bring a convention into existence or to choose the mode of ratification is outside Congress' constitutional authority."[272]

Copious evidence supports the conclusion that Congress may not refuse to call an amendments convention upon receiving the required number of applications.[273] When some Anti-Federalists suggested that Congress would not be required to call a convention,[274] Hamilton, writing in *Federalist No. 85* affirmed that the call would be mandatory.[275] Numerous other

---

271. William W. Van Alstyne, *Does Article V Restrict the States to Calling Unlimited Conventions Only?—A Letter to a Colleague*, 1978 DUKE L.J. 1295, 1303 (citing U.S. CONST. art. V).

272. Note, *Proposed Legislation on the Convention Method of Amending the United States Constitution*, 85 HARV. L. REV. 1612, 1633 (1972).

273. *See* CAPLAN, *supra* note 1, at 115–17.

274. *See, e.g.*, *Massachusettensis*, MASS. GAZETTE, Jan. 29, 1788, *reprinted in* 5 DOCUMENTARY HISTORY, *supra* note 1, at 831 (1998) ("Again, the constitution makes no consistent, adequate provision for amendments to be made to it by states, as states: not they who draught the amendments (should any be made) but they who ratify them, must be considered as making them. Three fourths of the legislatures of the several states, as they are now called, may ratify amendments, that is, if Congress see fit, but not without."); *A Customer*, N.Y.J., Nov. 23, 1787, *reprinted in* 19 DOCUMENTARY HISTORY, *supra* note 1, at 295 (2003) ("It is not stipulated that Congress shall, on the application of the legislatures of two thirds of the states, call a convention for proposing amendments.").

275. THE FEDERALIST NO. 85, *supra* note 1, at 456–57. Many writers have referenced this source, but few have discussed any of the corroborating sources discussed in this Part. *E.g.*, Ervin, *supra* note 1, at 885. THE FEDERALIST NO. 85 reads as follows:

> It is this that the national rulers, whenever nine States concur, will have no option upon the subject. By the fifth article of the plan, the Congress will be *obliged* "on the application of the legislatures of two thirds of the States, (which at present amount to nine) to call a convention for proposing amendments, which *shall be valid*, to all intents and purposes, as part of the Constitution, when ratified by the legislatures of three fourths of the States, or by conventions in three fourths thereof." The words of this article are peremptory. The Congress "*shall* call a convention." Nothing in this particular is left to discretion.

THE FEDERALIST NO. 85, *supra* note 1, at 456–57 (citing U.S. CONST. art. V).

Federalists agreed, among them James Iredell,[276] John Dickinson,[277] James Madison,[278] and Tench Coxe.[279] As Coxe observed:

> It has been asserted, that the new constitution, when ratified, would be fixed and permanent, and that no alterations or amendments, should those proposed appear on consideration ever so salutary, could afterwards be obtained. A candid consideration of the constitution will shew this to be a groundless remark. It is provided, in the clearest words, that Congress shall be *obliged* to call a convention on the application of two thirds of the legislatures. . . .[280]

---

276. 4 ELLIOT'S DEBATES, *supra* note 1, at 178 ("On such application, it is provided that Congress *shall* call such convention, so that they will have no option.").

277. *Fabius VIII*, PA. MERCURY, Apr. 29, 1788, *reprinted in* 17 DOCUMENTARY HISTORY, *supra* note 1, at 250 (1995) ("whatever their sentiments may be, they MUST call a Convention for proposing amendments, on applications of two-thirds of the legislatures of the several states").

278. Letter from James Madison to Thomas Mann Randolph (Jan. 19, 1789), *in* 11 THE PAPERS OF JAMES MADISON 415, 417 (Robert A. Rutland & Charles F. Hobson eds., 1977). Madison wrote: "It will not have escaped you, however, that the question concerning a General Convention, does not depend on the discretion of Congress. If two thirds of the States make application, Congress cannot refuse to call one; if not, Congress have no right to take the step." *Id.* at 417. Madison already had made the same point in another letter. *See* Letter from James Madison to George Eve (Jan. 2, 1789), *in* 11 THE PAPERS OF JAMES MADISON 4104, 405 (Robert A. Rutland & Charles F. Hobson eds., 1977).

279. Tench Coxe, *A Friend of Society and Liberty*, PA. GAZETTE, Jul. 23, 1788, *in* 18 DOCUMENTARY HISTORY, *supra* note 1, at 277, 283 (1995).

280. *Id.* at 283; *see also* Richard Law, Speech in the Connecticut Convention (Jan. 9, 1788), *in* 15 DOCUMENTARY HISTORY, *supra* note 1, at 316 (1984) ("a convention to be called at the instance of two thirds of the states"); *Solon, Jr.*, PROVIDENCE GAZETTE, Aug. 23, 1788, *reprinted in* 18 DOCUMENTARY HISTORY, *supra* note 1, at 340 (1995) ("But, secondly, although two-thirds of the New Congress should not be in favour of any amendments; yet if two-thirds of the Legislatures of the States they represent are for amendments, on the application of such two-thirds, the New Congress will call a General Convention for the purpose of considering and proposing amendments, to be ratified in the same manner as in case they had been proposed by the Congress themselves."). Similarly, the *Hudson Weekly Gazette* noted:

> It has been urged that the officers of the federal government will not part with power after they have got it; but those who make this remark really have not duly considered the constitution, for congress will be obliged to call a federal convention on the application of the legislatures of two thirds of the states: And all amendments proposed by such federal conventions are to be valid, when adopted by the legislatures or conventions of three fourths of the states. It therefore clearly appears that two thirds of the states can always procure a general convention for the purpose of amending the constitution, and that three fourths of them can introduce those amendments into the constitution, although the president, senate and federal house of representatives should be

Because of its agency role, Congress may—in fact, *must*—limit the subject matter of the convention to the extent specified by the applying states.[281] To see why this is so, consider an analogy: A property owner tells his property manager to hire a contractor to undertake certain work. The owner instructs the manager as to how much and what kind of work the contractor is to do. The manager is required to communicate those limits on the contractor and to enforce them.

In the state-application-and-convention procedure, the states are in the position of the property owner, Congress in the position of the manager, and the convention for proposing amendments in the place of the contractor. Historical evidence already adduced buttresses this conclusion,[282] showing that the applying state legislatures may impose subject-matter limits on the convention.

In order to carry out its agency responsibility, Congress has no choice, when counting applications toward the two thirds needed for a convention, but to group them according to subject matter.[283] Whenever two thirds of the states have applied based on the same general subject matter, Congress must issue the call for a convention related to that subject matter.[284] Congress may not expand the scope of the convention beyond that subject matter.[285] A recent commentary summarized the process this way:

> [A]pplications for a convention for different subjects should be counted separately. This would ensure that the intent of the States' applications is given proper effect. An application for an amendment addressing a particular issue, therefore, could not be used to call a convention that ends up proposing an amendment about a subject matter the state did not request be addressed. It follows from this argument that Congress's ministerial duty to call a convention also includes the duty to group applications according to subject matter. Once a sufficient number of applications have been reached, Congress must call a convention limited in scope to what the States have requested.[286]

---

unanimously opposed to each and all of them.

HUDSON WEEKLY GAZETTE, Jun. 17, 1788, *reprinted in* 21 DOCUMENTARY HISTORY, *supra* note 1, at 1200, 1201 (2005).

281. *See* Richard Law, *supra* note 280, at 316–17.

282. *See supra* Part IX.

283. CAPLAN, *supra* note 1, at 105.

284. James Kenneth Rogers, Note, *The Other Way to Amend the Constitution: The Article V Constitutional Convention Amendment Process*, 30 HARV. J.L. & PUB. POL'Y 1005, 1018 (2007).

285. CAPLAN, *supra* note 1, at 113.

286. Rogers, *supra* note 284, at 1018–19; *accord* Note, *Amendments*, *supra* note 1, at 1072; Kauper, *supra* note 1, at 911–12; Harmon, *supra* note 1, at 407 ("Unless there is general agreement among two-thirds of the legislatures over the nature of the change, or the

Of course, this is one area where "ministerial" duties necessarily require a certain amount of discretion, since Congress may have to decide whether differently worded applications actually address the same subject.[287]

C. Other Formalities in the Call

Article V bestows powers on named assemblies rather than on all actors in the legislative process.[288] That is why governors are excluded from the process.[289] This characteristic of Article V also suggests that the President has no role in calling a convention for proposing amendments—which is consistent with the earlier reference to the congressional role in the call as a procedural "throw-back" to pre-constitutional practice.[290]

The conclusion that the President has no role is buttressed both by a representation made by Federalist Tench Coxe during the ratification battle,[291] and by early ratification practice: Neither the congressional resolution forwarding the Bill of Rights to the states (1789) nor the resolution referring to them the Eleventh Amendment (1794) was presented to President Washington. Nor, apparently, did anyone suggest at the time that they should be.[292]

The Supreme Court has held that Congress may propose amendments by a two thirds vote of members present, assuming a quorum, not of the entire membership.[293] By parity of reasoning, Congress should be able to call the convention by majority of members present, assuming a quorum.

D. Enforcing the Duty to Call

The Constitution occasionally bestows authority of a kind normally exercised by one branch on another branch. The President is the chief executive, but he has a veto over bills, which is essentially a legislative power.[294] The Senate is usually a legislative body, but it enjoys power to try impeachments, a judicial power,[295] and to approve nominations, an

---

area where change is needed . . . the amendment process cannot go forward via the convention route.").

287. CAPLAN, *supra* note 1, at 105.

288. *Id.*

289. *Id.*

290. *Supra* Part X.A.

291. *A Friend of Society and Liberty*, PA. GAZETTE, July 23, 1788, *reprinted in* 18 DOCUMENTARY HISTORY, *supra* note 1, at 277, 283 (1995).

292. *Accord* CAPLAN, *supra* note 1, at 134–37; *see also* Hollingsworth v. Virginia, 3 U.S. (3 Dall.) 378 (1798) (holding that the President has no role in congressional amendment proposals).

293. Rhode Island v. Palmer, 253 U.S. 350, 350 (1920). This holding was foreshadowed by a similar decision in *Erkenbrecher v. Cox*, 257 F. 334, 336 (D. Ohio 1919).

294. U.S. CONST. art. I, § 7, cls. 2 & 3.

295. *Id.*, art. I, § 3, cl. 6.

executive power.[296] Congress serves as the federal legislature, but the Constitution grants it the power to declare war, which under the British Crown had been considered an executive power.[297]

In calling the convention, Congress wields an executive power. Because calling a convention is a mandatory executive duty, it should be enforceable judicially.[298] One potential remedy against a recalcitrant Congress is a declaratory judgment.[299] Because the duty is "plain, imperative, and entirely ministerial" a writ of mandamus also is appropriate.[300] In addition, if a legislature is violating the Constitution, courts may grant equitable relief, such as an injunction.[301]

E. The Composition of the Convention

From time to time, well-intended members of Congress have introduced legislation to govern the election and proceedings of any future convention for proposing amendments.[302] This legislation is justified as incidental to the congressional "call" power under the Necessary and Proper Clause.[303] Under some proposals, delegates would be allocated among the states by population or in proportion to their strength in Congress.[304]

Such legislation is constitutionally objectionable on several grounds. First, Founding-Era practice informs us clearly that choice over delegate

---

296. *Id.*, art. II, § 2, cl. 2.

297. *Id.*, art. I, §8, cl. 11; *see* NATELSON, ORIGINAL CONSTITUTION, *supra* note 1, at 124 (discussing the King's power to declare war).

298. *See* U.S. CONST. art. III, § 1, cls. 1.

299. Powell v. McCormick, 395 U.S. 486 (1969) (issuing a declaratory judgment retroactively reinstating an improperly evicted member of Congress).

300. Roberts v. United States, 176 U.S. 221, 230 (1900); *cf. McCormick*, 395 U.S. at 500–01 n.16, 517, 550 (not ruling out such relief against the relevant congressional officer). Rep. Theodore Sedgwick, an attorney speaking to the First Congress, noted the possibility of mandamus against Congress or the Senate. 1 ANNALS OF CONG. 544 (1789) (Joseph Gales & Seaton eds., 1834).

301. *E.g.*, Cooper v. Aaron, 358 U.S. 1 (1958) (rejecting a state's contention that its legislature and governor were not bound by federal court injunction).

302. *See, e.g.*, Ervin, *supra* note 1. Discussions of later bills are found in Diamond, *supra* note 1, at 113, 130–33, 137–38.

303. This has been the apparent justification of proposed congressional legislation. *See, e.g.*, Ervin, *supra* note 1; *see also* Kauper, *supra* note 1, at 906–07. For another claim of broad congressional power, see Charles L. Black, Jr., *The Proposed Amendment of Article V: A Threatened Disaster*, 72 YALE. L.J. 957, 964 (1963). The contrary position on this point was adopted in Gerald Gunther, *The Convention Method of Amending the United States Constitution*, 14 GA. L. REV. 1, 23–24 (1979). However, Professor Gunther, like most academics who addressed the issue in the 1960s and 1970s, opposed a convention.

304. Ervin, *supra* note 1, at 893; Kauper, *supra* note 1, at 909; *see also* Note, *Amendments*, *supra* note 1, at 1075–76 (supporting congressional legislation to that effect).

selection is an incident of the power of state legislatures, not of Congress.[305] In *intra-state* conventions, representation was apportioned roughly according to population,[306] but in *federal* conventions the caller requested states to send delegates of their own choosing. The states themselves were the participants.[307] They determined who the delegates were to be and how they would be chosen.[308]

The view that amendments conventions were assemblies of equal states persisted after the Constitution was ratified: They were referred to as "federal conventions" and "conventions of the states," rather than as conventions of the people.[309] For example, the 1789 Virginia application provided in part:

> [T]he Constitution hath presented an alternative, by admitting the submission to a *convention of the States*. . . . We do, therefore, in behalf of our constituents . . . make this application to Congress, that a convention be immediately called, of *deputies from the several States*, with full power to take into their consideration the defects of the Constitution that have been suggested by the State Conventions, and report such amendments thereto as they shall find best suited to promote our common interests, and secure to ourselves and our latest posterity, the great and unalienable rights of mankind.[310]

The 1789 New York application sent the same message:

> [W]e, the Legislature of the State of New York, do, in behalf of our constituents . . . make this application to the Congress, that a *Convention of Deputies from the several States* be called as early as possible, with full powers to take the said Constitution into their consideration, and to propose such amendments thereto, as they shall find best calculated to promote our common interests, and secure to ourselves and our latest posterity, the great and unalienable rights of mankind.[311]

This view was no mere hangover from the Founding Era, nor was it a rhetorical device to emphasize state sovereignty. Forty-two years later, the Supreme Court referred to a convention for proposing amendments as a "convention of the states."[312] This remained the standard phrase for decades.[313]

---

305. CAPLAN, *supra* note 1, at 119.

306. *Id.*

307. *E.g.*, 2 HOADLY, *supra* note 1, at 578 (reporting a resolution of the 1780 Philadelphia convention as "a meeting of the states").

308. *Id.*

309. Natelson, *First Century*, *supra* note 1.

310. Natelson, *Amending*, *supra* note 1, at 14.

311. *Id.*

312. Smith v. Union Bank, 30 U.S. 518, 528 (1831).

313. Natelson, *First Century*, *supra* note 1, at 10, 13–14.

This background compels the conclusion that the Article V convention is a creature—or, in the words of a former assistant United States Attorney General, the "servant"[314]—of the state legislatures, not of Congress, nor of the people directly.[315] Those legislatures, therefore, determine how delegates are allocated and selected.[316]

Another problem with schemes by which Congress prescribes delegate selection procedures is that they undercut the congressional-bypass goal of the state-application-and-convention process.[317] The process would not be an effective bypass if Congress could set—or gerrymander—the convention's composition or rules.[318] Moreover, apportioning delegates in a way that does not treat all states equally violates Congress's fiduciary duty to treat impartially all states, who are its principals in this limited context.[319] How delegates are to be selected, or how many to send, is for principals, not agents, to decide.

F. Convention Discretion: The Rules

Under the incidental powers conferred by Article V, an amendments convention adopts its own rules and elects its own officers.[320] This follows from Founding-Era custom: All conventions, inter- or intra-state, established their own rules, judged their own credentials, carried out their own housekeeping, and elected their own officers.[321] Thus, the fixing of rules is not a matter either for Congress[322] or the applying states. More

---

314. Harmon, *supra* note 1, at 409.

315. *Cf.* EVERETT SOMERVILLE BROWN, RATIFICATION OF THE TWENTY-FIRST AMENDMENT TO THE CONSTITUTION OF THE UNITED STATES 516–17 (1938) (showing that on the one occasion when Congress opted for a proposed constitutional amendment to be ratified by state conventions rather than state legislatures, the states were left in full command of delegate-selection).

316. *Id.*

317. *Cf.* Diamond, *supra* note 1, at 144–45 (expressing approval of the idea of electing delegates by population, but affirming that it is beyond Congress's power to mandate this).

318. *Id.*

319. *See generally* Natelson, *Judicial Review*, *supra* note 1, at 262–267 (describing how fiduciaries are to treat their beneficiaries impartially).

320. U.S. CONST. art. V.

321. *See, e.g.*, 1 HOADLY, *supra* note 1, at 589 (reporting that the first Providence Convention was electing its officers); *id.* at 611 (reporting that the New Haven Convention was adhering to "one state, one vote"); 2 HOADLY, *supra* note 1, at 577 (reporting that the 1780 Philadelphia convention was choosing its own president and fixing a succession rule); 3 HOADLY, *supra* note 1, at 561 (reporting that the Boston Convention was electing its own officers); *id.* at 575 (reporting that the second Providence Convention was electing its own officers); 1 FARRAND'S RECORDS, *supra* note 1, at 7–9 (reporting that the 1787 Philadelphia convention was adopting its own rules); 3 ELLIOT'S DEBATES, *supra* note 1, at 3 (reporting that the Virginia ratifying convention was adopting its own rules).

322. The Ervin legislation included provisions for congressional governance. These

recently, the principle that a convention, or a legislature, operating under Article V controls its own rules and procedures, including voting rules, was applied by Justice Stevens in his much-quoted opinion in *Dyer v. Blair*.[323]

Suffrage is decided by convention rule. The convention is free to adjust its rules of suffrage however it wishes, but the initial suffrage rule is "one state, one vote."[324] This may seem undemocratic, but of course the Constitution erected a mixed federal government, not a purely democratic one.

The democratic interest is protected by Congress's ability to propose amendments, and also by the requirement that three fourths ratify a proposal for it to be effective.[325] Although it is possible theoretically for three fourths of the states to represent only a minority of the population,[326] it is nearly impossible as a matter of practical politics because of sharp differences in the political character among states of similar sizes.[327]

---

were supported by some writers based on views unshaped by the actual ratification record. *See, e.g.*, Kauper, *supra* note 1, at 909 (suggesting that Congress could require that delegates be elected by population). Based on a fuller review of the record, Caplan reaches substantially the same conclusions as I do. CAPLAN, *supra* note 1, at 119–20.

323. Dyer v. Blair, 390 F. Supp. 1291, 1307 (N.D. Ill. 1975) ("Article V identifies the body—either a legislature or a convention—which must ratify a proposed amendment. The act of ratification is an expression of consent to the amendment by that body. By what means that body shall decide to consent or not to consent is a matter for that body to determine for itself."). Although Justice Stevens was referring to a ratifying body, there is no reason this rule should not apply to an amendments convention.

324. *See, e.g.*, 1 HOADLY, *supra* note 1, at 611 (reporting that the New Haven Convention was adhering to "one state, one vote"). This follows from the treatment of delegations as units, i.e., as "committees." *See supra* note 75 and accompanying text. If a state opted for district elections for delegates, the Equal Protection Clause of the Fourteenth Amendment, which the United States Supreme Court has construed as containing a "one person one vote rule," would apply within the state. CAPLAN, *supra* note 1, at 120. That rule should have no effect, however, at the federal level, when states act, either directly or through a convention, as states. One appropriate analogy is the United States Senate; a closer one is the ratification of constitutional amendments by three-fourths of the states, irrespective of population.

325. U.S. CONST. art. V.

326. UNITED STATES CENSUS BUREAU, ANNUAL ESTIMATES OF THE POPULATION FOR THE UNITED STATES, REGIONS, AND STATES AND FOR PUERTO RICO (2006). According to United States Census Bureau 2006 population estimates, if *all* the twelve largest states opposed ratification and *all* the thirty-eight smallest ratified, then the ratifying states would contain only a little more than forty percent of the American people. This scenario would require unanimity among the twelve largest states, which are quite disparate politically, and unanimity among the thirty-eight smallest, which are similarly diverse. The first group includes such disparate pairs as Massachusetts and Texas, New York and North Carolina, and Michigan and Georgia. The second group includes states such as Hawaii and Wyoming, Vermont and Colorado.

327. Kauper, *supra* note 1, at 914, pointed this out in 1966, and state population disparities were slightly *greater* then than they now are.

Approval by three fourths of the states will reflect majority, and probably super-majority, public support.[328]

G. Convention Discretion: An Application May Not Limit the Convention to Specific Rules or Language

Some comparatively recent applications have tried to impose restrictions beyond subject-matter limits. For example, some have purported to require the convention to take an up-or-down vote on an amendment whose precise wording is set forth in the application.[329] Applications also have imposed conditions precedent to operation (providing that the application becomes effective only when a certain event or events occur)[330] and conditions subsequent (providing that the application becomes ineffective if a particular event or events intervene).[331] Some applications have included both kinds of conditions.[332]

These restrictions were imposed to guard against the supposed danger of a "runaway" convention, but what they really do is create practical and legal problems. The *practical* problems arise from the fact that the more terms and conditions applications contain, the less likely they will match each other sufficiently to be aggregated together to reach the two-thirds threshold.[333] Members of Congress and judges who dislike the contemplated amendments may seize upon wording differences to justify refusal to aggregate.[334]

The *legal* difficulties arise because the courts are likely to reject any effort by state legislatures to impose rules or specific language on the convention. The universal prerogative of conventions during the Founding Era[335] and after[336] has been to make their own rules, and in modern times

---

328. UNITED STATES CENSUS BUREAU, *supra* note 326.

329. *E.g.*, 133 CONG. REC. 7299 (Mar. 30, 1987) (reproducing Utah application specifying precise text of amendment).

330. CONG. GLOBE, 36TH CONG., 2ND SESS. 680 (Feb. 1, 1861) ("[U]nless the remedies before suggested be speedily adopted, then, as a last resort, the State of New Jersey hereby makes application, according to the terms of the Constitution, of the Congress of the United States, to call a convention (of the States) to propose amendments...").

331. 133 CONG. REC. 7299 (Utah application stating that it becomes void if Congress proposes an identical amendment).

332. *E.g.*, 139 CONG. REC. 14,565 (Jun. 29, 1993) (Missouri application containing condition precedent of congressional non-action, followed by condition subsequent of congressional action).

333. *See generally supra* note 329 (Utah application specifies precise text of the amendment to be adopted).

334. CAPLAN, *supra* note 1, at 107–08, suggests that refusal to aggregate would be improper, and that applications could be amended to comply with each other.

335. *E.g.*, 1 FARRAND'S RECORDS, *supra* note 1, at 7–9, 14–16 (discussion and agreement to rules of Constitutional Convention); 2 ELLIOT'S DEBATES, *supra* note 1, at 1 (appointment of rules committee at Massachusetts ratifying convention); 3 ELLIOT'S

the courts have defended the power of Article V assemblies to do so.[337] Courts also have defended the power of Article V assemblies to deliberate and to exercise discretion.[338] Opponents may argue that if an application purports to prescribe rules or specific language to the convention, it is void for attempting to obtain an illegal result.[339]

One purpose of the state-application-and-convention process was to give state legislatures a role nearly co-equal to Congress as a promoter of amendments. Allowing states to dictate rules and language in their applications arguably serves that purpose. But a competing purpose was to ensure that the actual proposals come from a single deliberative body representing all, not only the applying, state legislatures.[340] The text of the Constitution grants the *convention*, not the state legislatures, the ultimate power to "propos[e] Amendments."[341] The Framers could have drafted language permitting the states to propose amendments directly, but they did not.

The Framers inserted a convention into the process presumably because the convention setting encourages collective deliberation, compromise, and conciliation. Deliberation requires the ability to weigh alternatives or even, as Madison and others suggested during the ratification fight, the power not to propose at all.[342]

DEBATES, *supra* note 1, at 3 (recording Virginia ratifying convention as adopting rules of state House of Delegates); PROCEEDINGS OF THE CONVENTIONS OF THE PROVINCE OF MARYLAND 3 (Baltimore, James Lucas & E.K. Deaver eds., 1836) (reporting that the 1774 provincial convention adopted its own voting rule).

336. HOAR, *supra* note 1, at 170–84 (discussing the rule-making power of conventions).

337. *E.g.*, Dyer v. Blair, 390 F. Supp. 1291 (N.D. Ill. 1975).

338. *See infra* notes 347–354 and accompanying text.

339. *Cf.* Arthur E. Bonfield, *The Dirksen Amendment and the Article V Convention Process*, 66 MICH. L. REV. 949, 959 (1968) (arguing that applications seeking ratification by state legislatures rather than state convention seek an illegitimate end and should be disregarded).

340. *Cf.* Dodge v. Woolsey, 59 U.S. 331, 348 (1855). In *Dodge*, the Court stated of the amendment process that

> [T]he people of the United States, aggregately and in their separate sovereignties . . . have excluded themselves from any direct or immediate agency in making amendments to [the Constitution], and have directed that amendments should be made representatively for them, by the congress of the United States, when two thirds of both houses shall propose them; or where the legislatures of two thirds of the several States shall call a convention for proposing amendments [subject to state ratification].

Id. at 348. The implication is that the states, the people's "separate sovereignties," cannot dictate directly amendments themselves, and that the drafting and proposal are the prerogatives of Congress or the convention.

341. U.S. CONST. art. V.

342. James Madison to Philip Mazzei (Dec. 10, 1788), *in* 11 THE PAPERS OF JAMES

Admittedly, a large number of applications with similar restrictions also are likely to be the product of considerable deliberation and some compromise and conciliation.[343] But the convention setting encourages more, and includes the non-applying states. An independent level between state applications and state ratification subjects the process of decision to being further "refined," to use Madison's term.[344]

History paints a picture of what the Founders had in mind. Founding-Era interstate conventions could be—and usually were—limited to particular subject matter.[345] Yet they invariably were deliberative entities, if not always among delegates, then at least among state delegations.[346] No one imposed "take it or leave it" language in the call.[347] The conventions proposed; and as incidents to their power to propose, they deliberated and drafted.[348] As noted earlier,[349] the resulting procedure closely parallels how the first ten amendments were adopted: First, the states suggested a number of amendments.[350] Then, working almost entirely from that list, Congress (here, acting much as an amendments convention would) deliberated the merits of each, selected some of the states' ideas, performed the actual drafting, and sent its proposals back to the states for ratification.[351]

This is another topic on which most subsequent history is consistent with the Founders' vision. Throughout the nineteenth and early twentieth centuries, no application, even an application limited to a particular subject matter, sought to dictate precise wording or terms to the convention.[352] At least one application was subject to a condition: An 1861 New Jersey application was to be effective only if Congress did not act.[353] But that

---

MADISON 1788–1789, at 389 (Robert A. Rutland & Charles F. Hobson eds., 1977); *see also* Letter from An Old Whig II, PHILA. INDEP. GAZETTEER (Oct. 12, 1787), *reprinted in* 13 DOCUMENTARY HISTORY, *supra* note 1, at 376, 377 (1981) (observing, shortly after the Constitution became public, "the convention may agree to the [states-suggested] amendments or not as they think right").

343. CAPLAN, *supra* note 1, at 105.

344. *See, e.g.*, THE FEDERALIST NO. 10, *supra* note 1, at 46 (asserting that when a decision is passed through a chosen body of citizens the effect is to "refine and enlarge the public views").

345. *See supra* Part IX.A..

346. *See* Hawke v. Smith, 253 U.S. 221, 226–27 (1920).

347. *See generally Convention Applications*, *supra* note 78.

348. *See Hawke*, 253 U.S. at 226–27.

349. *See supra* note 249 and accompanying text.

350. *See supra* note 251 and accompanying text.

351. *See supra* note 252 and accompanying text.

352. *See Convention Applications*, *supra* note 78.

353. CONG. GLOBE, 36TH CONG., 2ND SESS. 680 (Feb. 1, 1861) ("unless the remedies before suggested be speedily adopted, then, as a last resort, the State of New Jersey hereby makes application, according to the terms of the Constitution, of the Congress of the United States, to call a convention (of the States) to propose amendments").

condition did not infringe the assembly's deliberative freedom once the convention had been called.[354]

In the 1930s, state legislatures explored ways to restrict the deliberative freedom of Article V assemblies by assuring adherence to the popular will.[355] This effort won judicial approval in the 1933 Alabama Supreme Court advisory opinion, *In re Opinion of the Justices*.[356] The issue was a state law governing the convention called for ratifying or rejecting the Twenty-First Amendment repealing Prohibition.[357] The statute provided that an elector's vote for convention delegates would not be counted unless the elector first voted "yes" or "no" on the question of whether Prohibition should be repealed.[358] The law required delegates to take an oath promising to support the result of the referendum.[359] The court sustained this procedure as promoting the popular will.[360] The court gave little or no weight to the goal of assuring a deliberative process.[361]

However, if Assembly X effectively restricts the deliberation of Assembly Y, some of Assembly Y's decision-making authority is transferred to Assembly X. By absolutely binding the convention to the popular will, the Alabama statute effectively transferred ratification from the convention to the voters.[362] They became the true ratifiers.[363] For this reason, other courts have not followed *In re Opinion of the Justices*.[364]

Even before that case, the Supreme Court had decided that a ratifying assembly could not be displaced by a referendum[365] and that an assembly's discretion could not be compromised by extraneous rules.[366] In the same year as *In re Opinion of the Justices*, the Supreme Court of Maine ruled that a referendum cannot bind a ratifying convention because "[t]he convention must be free to exercise the essential and characteristic function of rational deliberation."[367]

Since that time, a string of holdings has recognized explicitly the connection between control and deliberation, and has done so in the application context as well as in ratification context. In 1978 Justice

---

354. *See id.*

355. *See In re* Opinion of the Justices, 148 So. 107 (Ala. 1933).

356. *Id.* at 111.

357. *See id.* at 108.

358. *Id.*

359. *Id.*

360. *Id.* at 110.

361. *See generally id.* at 110–11.

362. *See id.*

363. *See id.*

364. *See, e.g.*, State *ex rel.* Harper v. Waltermire, 691 P.2d 826 (Mont. 1984); AFL-CIO v. Eu, 686 P.2d 609 (Cal. 1984), *stay denied sub nom.* Uhler v. AFL-CIO, 468 U.S. 1310 (1984) (advisory resolution).

365. Hawke v. Smith, 253 U.S. 221 (1920).

366. Leser v. Garnett, 258 U.S. 130, 137 (1922) (citations omitted).

367. *In re* Opinion of the Justices, 167 A. 176, 180 (Me. 1933).

Rehnquist upheld a referendum to influence the application process, but emphasized that the referendum was purely advisory.[368] Six years later, the Montana Supreme Court voided an initiative that would have required state lawmakers to apply for a convention for proposing a balanced budget amendment.[369] Relying on the United States Supreme Court cases disallowing transfer of ratifying power to the voters, the Montana tribunal held that, "[a] legislature making an application to Congress for a constitutional convention under Article V must be a freely deliberating representative body. The deliberative process must be unfettered by any limitations imposed by the people of the state."[370]

The same year, the California Supreme Court invalidated a voter initiative imposing financial penalties on lawmakers who failed to support an application for a balanced budget amendment.[371] The court observed that this was inconsistent with a goal of Article V, which "envisions legislators free to vote their best judgment."[372]

During the 1990s battle for federal term limits, activists used the state initiative process to induce lawmakers to support their cause.[373] Members of Congress were instructed to support congressional proposal of a term limits amendment.[374] State lawmakers were instructed to support applications for a convention that would propose term limits.[375] Voter-adopted initiatives inflicted negative ballot language on politicians who refused.[376] Again and again courts invalidated these measures, because by impeding the deliberative function they transferred discretion from Article V assemblies to other actors.[377] Although one could interpret those measures as a form of

---

368. *Kimble* v. Swackhamer, 439 U.S. 1385, *appeal dismissed*, 439 U.S. 1041 (1978) (Rehnquist, J.); *see also AFL-CIO*, 686 P.2d 609.

369. State *ex rel.* Harper, 691 P.2d 826.

370. *Id.* at 830 (citing *Leser*, 258 U.S. 130).

371. *See AFL-CIO*, 686 P.2d 609.

372. *Id.* at 613.

373. *See, e.g.*, Miller v. Moore, 169 F.3d 1119 (8th Cir. 1999); Gralike v. Cook, 191 F.3d 911, 924–25 (8th Cir. 1999), *aff'd on other grounds sub nom.* Cook v. Gralike, 531 U.S. 510 (2001); Barker v. Hazetine, 3 F. Supp. 2d 1088, 1094 (D.S.D. 1998); League of Women Voters of Maine v. Gwadosky, 966 F. Supp. 52 (D. Me. 1997); Donovan v. Priest, 931 S.W.2d 119 (Ark. 1996).

374. *See, e.g.*, *Miller*, 169 F.3d at 1121–22.

375. *See Gralike*, 191 F.3d at 925.

376. *See id.* (citations omitted).

377. *E.g.*, *Miller*, 169 F.3d 1119; *Gralike*, 191 F.3d at 924–25 ("Article V envisions legislatures acting as freely deliberative bodies in the amendment process and resists any attempt by the people of a state to restrict the legislatures' actions."); *Barker*, 3 F. Supp. 2d at 1094 ("Without doubt, Initiated Measure 1 brings to bear an undue influence on South Dakota's congressional candidates, and the deliberative and independent amendment process envisioned by the Framers when they drafted Article V is lost."); *League of Women Voters of Maine*, 966 F. Supp. 52; *Donovan*, 931 S.W.2d at 127, (requiring an assembly that can engage in "intellectual debate, deliberation, or consideration").

aggressive advice rather than actual coercion, the courts consistently invalidated them.[378]

As an application campaign nears apparent success, it will be opposed by hostile opinion makers, judges, and members of Congress.[379] They will contend that applications restricting convention discretion are inherently void.[380] As to the specification of subject matter, there is ample response: the kind of convention the Founders had in mind was the task force assigned one or more subjects to address.[381] It also is clear that legislatures may make *recommendations* in their applications.[382] Legislatures that go much further place their applications at risk.

H. State Legislative Instructions

The deliberative quality of the convention does not mean that the delegates are, within the topic of the convention, completely free actors. American convention delegates have long been subject to instructions from those they represent.[383] As in all prior federal conventions, delegates to a convention for proposing amendments are representatives of the state legislatures, and therefore subject to instructions.[384]

This is suggested also by Madison's comment in *Federalist No. 43* that Article V "equally enables the general and the state governments, to originate the amendment of errors . . . ."[385] Since Congress may propose amendments directly to the states for ratification or rejection, granting equal (or nearly equal) power to the states requires either that they can propose directly (which they cannot) or that they act through convention delegates who are their agents. There is no third alternative.[386]

The power to instruct by no means precludes deliberation. Delegates can discuss and negotiate issues among themselves and with the home office. The home office can discuss and negotiate with their counterparts in other states. The result will be a textured, multi-layered deliberation likely superior to anything that the delegates could have produced alone.

---

378. *See supra* note 373.

379. *See* Black, *supra* note 1.

380. *See, e.g., id.* at 190–92 (arguing that an application referencing specific language should be disregarded).

381. *See* Natelson, *Amending*, *supra* note 1, at 5–9; Natelson, *First Century*, *supra* note 1, and discussion above.

382. The state ratifying conventions made extensive recommendations for amendments to be acted on either by Congress or by an Article V convention. *See also* Kimble v. Swackhamer, 439 U.S. 1385 (1978); AFL-CIO v. Eu, 686 P.2d 609 (Cal. 1984).

383. HOAR, *supra* note 1, at 127–29.

384. *See id.*

385. THE FEDERALIST NO. 43, *supra* note 1, at 228.

386. *See* Hawke v. Smith, 253 U.S. 221, 227 (1920); THE FEDERALIST NO. 43, *supra* note 1, at 228.

XI. RULES GOVERNING TRANSMITTAL OF PROPOSALS TO THE STATES

A. What Happens if the Convention "Proposes" an Amendment Outside the Subject Assigned by the Applications?

Because the convention serves the state legislatures, only proposals within the subject matter fixed by the applications, and therefore within the convention call, have legal force. Actions outside the call are *ultra vires* and legally void. Yet under agency law, both at the Founding and today, an agent may suggest to his principal a course of action outside the agent's sphere of authority. This suggestion, however valuable, is a recommendation only, without legal force. For example, if a convention called to consider a balanced budget amendment recommends both a balanced budget amendment and a term limits amendment, only the former is a "proposal" within the meaning of Article V.[387] The latter is merely a recommendation for future consideration.[388] In the words of President Carter's Assistant Attorney General John Harmon, the convention delegates "have . . . no power to issue ratifiable proposals except to the extent that they honor their commission."[389]

Thus, Congress may specify a "Mode of Ratification" only for proposals within the convention call, and states may ratify only proposals within the call.[390] If Congress, the legislatures, or the public agrees with the convention's *ultra vires* recommendation, the states may apply anew for a convention with authority to propose them or Congress itself may propose them.[391]

B. Choosing a Mode of Ratification

Although a convention's proposal does not technically pass through Congress to the states, the Constitution does require and empower Congress to select one of two "Modes of Ratification."[392] Congress's power in this regard is the same as if it had proposed the amendment.[393] Article V alters the normally subservient position to the states that Congress usually occupies in the state-application-and-convention process[394] by prescribing

---

387. *See* CAPLAN, *supra* note 1, at 147, 157.

388. *See id.*

389. Harmon, *supra* note 1, at 410.

390. *See* CAPLAN, *supra* note 1, at 147.

391. Natelson, *Amending*, *supra* note 1, at 15.

392. *See* CAPLAN, *supra* note 1, at 147.

393. *See id.*

394. That this is a departure from the normal state-driven process is underscored by the fact that state-power advocate Elbridge Gerry moved during the federal convention to strike it. The convention refused:

that Congress, not the state legislatures, will decide on whether ratification is by state legislatures or by state conventions.[395]

However, Congress has no choice as to whether to choose a "Mode."[396] The Constitution requires it to do so.[397] Because selecting, like calling an Article V convention, is a mandatory rather than discretionary duty, it should be enforceable judicially.[398] On the other hand, congressional discretion as to choice of method is unreviewable.[399]

Congress may enjoy some powers incidental to the power to select a mode of ratification, but if so, they are quite circumscribed. As we have seen, under the doctrine of incidental authority incorporated into Article V, Power B may not be incidental to Power A if Power B is as great or greater than Power A, or if not coupled with it by custom or strong necessity.[400] The power to choose the mode of ratification is obviously a limited and discrete one, and certainly does not justify sprawling congressional authority over the state ratification process. The Supreme Court's holding in *Dillon v. Gloss*[401]—that Congress may specify a time period for

---

> Mr [sic] Gerry moved to strike out the words "or by Conventions in three fourths thereof"
> On this motion
> N-- H-- no. Mas. no-- Ct. ay. N-- J. no. Pa no--Del-- no. Md no. Va. no. N-- C. no. S. C. no-- Geo-- no. [Ayes -- 1; noes -- 10.]
> Mr. Sherman then moved to strike out art V altogether
> Mr [sic] Brearley 2ded. the motion, on which
> N. H. no. Mas. no. Ct. ay. N. J. ay. Pa. no. Del. divd. Md. no. Va. no. N. C. no. S. C. no. Geo. no [Ayes -- 2; noes -- 8; divided -- 1.]

2 FARRAND'S RECORDS, *supra* note 1, at 630–31.

395. *See* CAPLAN, *supra* note 1, at 147.

396. *See id.*

397. *See* U.S. CONST. art. V.

398. *See supra* notes 377–378 and accompanying text. Note, however, that during the ratification fight, two Anti-Federalists argued that Congress could sabotage the state-application-and-convention process by failing to transmit the convention's proposed amendments to the states. "Samuel," INDEPENDENT CHRONICLE, Jan. 10, 1788, *reprinted in* 5 DOCUMENTARY HISTORY, *supra* note 1, at 678, 682 (1998) ("Moreover, could we obtain a Convention, and by them amendments proposed; they might lie dormant forever, if the Congress did not see cause to appoint how the amendments should be ratified; which is not to be expected, if the amendments should be to diminish their power"); Letter from An Old Whig VIII, PHILA. INDEP. GAZETTEER (Feb. 6, 1788), *reprinted in* 16 DOCUMENTARY HISTORY, *supra* note 1, at 52–53 (2001) ("such amendments afterwards to be valid if ratified by the legislatures of three-fourths of the states, or by conventions in three-fourths thereof, if Congress should think proper to call them"). Such a construction would, of course, undercut the fundamental purpose of the state-application-and-convention process, and should be disfavored if only for that reason.

399. *See* U.S. CONST. art. V.

400. *See supra* notes 40–49 and accompanying text.

401. 256 U.S. 368 (1920).

ratification as an incident of selecting the mode—may or may not be correct, but it certainly should apply only when the proposal comes from Congress. Congress may specify a time period for *its own* proposed amendments, since proposers generally may impose time limits on their own proposals. But when a convention proposes amendments, the convention, not Congress, is the correct agency for setting the time limit. Vesting the power in Congress would be inconsistent with the purpose of the state-application-and-amendment process, since it would enable Congress to throttle proposals it dislikes by imposing very short time limits.[402]

XII. CONCLUSION

Because a convention for proposing amendments has never been called, the state-application-and-convention process seems mysterious to some. Convention opponents have taken advantage of the mystery by summoning specters of their own devising.

There need be little mystery. The nature of the process is recoverable from American history and American law. This paper explains the principal customs of interstate conventions during the Founding and how they illuminate the Article V process. It explains why the Founders included the process in the Constitution, and how they expected it to operate. It draws on nearly two centuries of experience and case law that are generally consistent with the Founders' design. While this paper does not answer all questions, it does answer some fundamental ones.

The issues that remain will be resolved as state lawmakers and other citizens invoke the process. Those issues will be resolved by mutual consultation and, perhaps in a few instances, by judicial decision. There is nothing unusual in this: As the Founders recognized, some constitutional questions can be elucidated only through practice.[403] If they had insisted that every question be answered in advance, they never would have bequeathed to us either the Constitution or the Bill of Rights.

Refraining from the state-application-and-convention process is not honoring the Constitution. Quite the contrary: Because the process was inserted in the document for what the Framers and Ratifiers considered very compelling reasons, ignoring it leaves the instrument incomplete—indeed, may cripple it. Without a vigorous state-application-and-convention process, the Constitution's checks and balances are not fully effective after all.

---

402. Rees, *supra* note 1, at 93–94.

403. THE FEDERALIST NO. 82, *supra* note 1, at 426 ("Time only can mature and perfect so compound a system, liquidate the meaning of all the parts, and adjust them to each other in a harmonious and consistent WHOLE.").

SOVEREIGNTY, REBALANCED: THE TEA PARTY & CONSTITUTIONAL AMENDMENTS

ELIZABETH PRICE FOLEY*

"[E]xperience hath shewn, that even under the best forms, those entrusted with power have, in time, and by slow operations, perverted it into tyranny. . . ." –Thomas Jefferson[1]

Jefferson's words ring true today. Arguably since the Marshall Court and undoubtedly since the New Deal, the U.S. Constitution has been subverted to the point where its original meaning has been substantially lost inside a tangled knot of Supreme Court case law. Like termites eating away at the constitutional architecture, Supreme Court interpretations of provisions such as the Commerce Clause, taxing and spending power, Privileges or Immunities Clause, Ninth Amendment, Tenth Amendment, and Eleventh Amendment have so rotted them that they no longer serve the critical functions originally envisioned.

One of the most pervasive themes in this journey into constitutional Wonderland—where constitutional law professors teach at least six impossible things before breakfast—is the loss of vertical separation of powers, or federalism. Year after year, the drumbeat of expanding federal power grows louder, drowning out objections and concerns voiced by the states. The noise has recently reached a fevered pitch, fueled by actions of the Obama Administration: massive industry bailouts overloaded with federal strings, mind-numbing trillion-dollar stimulus programs laden with earmarks, aggressive use of federal powers to shut down states' efforts to fight illegal immigration, and, the *coup de grâce*, Obamacare.[2]

The six-million-dollar question is how to untangle this constitutional Gordian knot. The most intriguing proposals call for specific constitutional amendments or a constitutional convention, the latter of which has not occurred since the grand convention in Philadelphia that fateful, hot

---

* Institute for Justice Chair in Constitutional Litigation and Professor of Law, Florida International University College of Law. Professor Foley is the author of LIBERTY FOR ALL: RECLAIMING INDIVIDUAL PRIVACY IN A NEW ERA OF PUBLIC MORALITY (2006) and THE LAW OF LIFE AND DEATH (2011). She is currently working on a book about the Tea Party that will be published by Cambridge University Press in early 2012.

1. Thomas Jefferson, A Bill for the More General Diffusion of Knowledge, *in* 2 THE WORKS OF THOMAS JEFFERSON 414 (Paul Leicester Ford ed., 1904).

2. *See generally* THE HERITAGE FOUND., *Federal Budget and Spending*, ISSUES 2010: THE CANDIDATE'S BRIEFING BOOK 39–43 (2010), http://issues2010.com/pdf/Full_PDF.pdf; Jackie Calmes, *Obama to Seek Spending Freeze to Trim Deficits*, N.Y. TIMES, Jan. 26, 2010, at 1.

summer of 1787.[3] This essay will explore the major themes of these calls for constitutional amendments and conventions, who is behind them, what problems they seek to solve, and their likelihood of success.

I. WHY REBALANCING IS NEEDED

A refrain commonly encountered when discussing federalism is something like this: "Who cares what a bunch of dead prejudiced white guys thought about states' rights? The Civil War was fought in the name of states' rights and the South lost. We should care more about what modern society needs from government than about turning back the constitutional clock in the name of some outdated federalism fetish."[4] Professor Michael Klarman sums up this attitude with the pejorative label, "constitutional idolatry."[5] Buried not too deeply behind this label is a liberal-progressive ideology harboring a deep-seated fear that federalism is secret code for supporting slavery and segregation and opposing things like abortion, gay marriage and, most recently, health care reform.

What these Constitution vilifiers fail to grasp is that modern Americans who decry the erosion of federalism are not pining for a return to segregation or some pre-Civil War version of states' rights. Instead, they want to maximize individual liberty by identifying and enforcing meaningful limits on federal power. Put another way, federalism proponents are not romanticizing a bygone era. They are trying to preserve a constitutional principle—a vigorous system of dual sovereignty—that is designed to compete for the affection of citizens and jealously guard their rights. In the words of James Madison in *Federalist No. 45*:

> Several important considerations have been touched in the course of these papers, which discountenance the supposition that the operation of the federal government will by degrees prove fatal to the State governments. The more I revolve the subject, the more fully I am persuaded that the balance is much more likely to be disturbed by the preponderancy of the last than of the first scale.

---

3. *See* Gerald Gunther, *The Convention Method of Amending the United States Constitution*, 14 GA. L. REV. 1, 2 (1979).

4. *See, e.g.*, *The Perils of Constitution-Worship*, THE ECONOMIST, Sept. 25, 2010, at 46. (Describing conservatives' and tea partiers' emphasis on the Constitution and Declaration of Independence as indicative of the "same dream of return to prelapsarian innocence" unwarranted because the Framers were "aristocrats, creatures of their time fearful of what they considered the excessive democracy taking hold in the states in the 1780s. They did not believe that poor men, or any women, let alone slaves, should have the vote.").

5. Michael Klarman, *A Skeptical View of Constitution Worship*, BALKINIZATION (Sept. 16, 2010, 6:34 AM), http://balkin.blogspot.com/2010/09/skeptical-view-of-constitution-worship.html.

> . . . [T]he States will retain, under the proposed Constitution, a very extensive portion of active sovereignty. . . .
>
> The State government will have the advantage of the Federal government, whether we compare them in respect to the immediate dependence of the one on the other; to the weight of personal influence which each side will possess; to the powers respectively vested in them; to the predilection and probable support of the people; to the disposition and faculty of resisting and frustrating the measures of each other.
>
> The State governments may be regarded as constituent and essential parts of the federal government; whilst the latter is nowise essential to the operation or organization of the former.[6]

Madison could not be any clearer in his message to the American people who ratified the Constitution: state sovereignty not only exists, but it exists for the benefit of "We the People." So it is both extremely simplistic and borderline disingenuous to suggest that federalism proponents are motivated by a desire to protect states qua states. The motivation is to protect "We the People," and federalism is a critically important structural mechanism for doing this.

Even assuming you are convinced that federalism is more than just a quaint antediluvian relic, the question arises as to *why* so many people suddenly seem to think it needs to be restored. The Civil War Amendments undeniably shifted power away from the states.[7] The New Deal Court's Commerce Clause jurisprudence aggrandized federal power at the expense of the states. But these seismic shifts occurred long ago. So why is an audible cry of "federalism!" only now emerging, in *Horton Hears a Who!* fashion, from a seemingly small speck of intellectual dust?

There is no single reason. Extant angst over federalism is based not only on the collective impact of Supreme Court decisions regarding commerce, the spending power, the Tenth Amendment, etc., but on several recent events that seem to have broken the proverbial camel's back. Since late 2008, the federal government has been spending like a drunken sailor on leave, with no apparent awareness of the responsibility to repay its debt. It responded to a free-falling economy by spending trillions of dollars to bail out banks, brokerage houses, automakers, Fannie Mae and Freddie Mac, pension funds and others. Additional massive stimulus laws have doled out hundreds of billions more for infrastructure projects, unemployment and food-stamp benefits, shoring up state education and Medicaid, and various congressional pet projects.[8]

On top of all this, in the face of some of the most uncertain economic times ever experienced, the Obama Administration used strong-arm tactics

---

6. THE FEDERALIST NO. 45 (James Madison) (George F. Hopkins ed., 1802).

7. *See generally* John E. Nowak, *Federalism and the Civil War Amendments*, 23 OHIO N.U. L. REV. 1209 (1997).

8. *See generally* John B. Taylor, *The Coming Debt Debacle: Top Economist Says President Obama Must Slash Spending, Now*, N.Y. DAILY NEWS, Aug. 31, 2009.

to pass a nearly trillion-dollar health reform bill over the objections of most Americans.[9] Obamacare purports to rely upon Congress's power to regulate interstate commerce as its basis for imposing heavy regulations on the health insurance industry—an area traditionally regulated by the states—and mandating that individuals buy private health insurance.[10] This unprecedented assertion of federal power, if sustained, would effectively give the federal government a police power, affecting individual liberty in an unprecedented and limitless way.

The net result of the federal government's intoxicated behavior has been a federal power grab of a magnitude never seen before. Americans have become disgusted with the behavior of the President and Congress, viewing their behavior as motivated more by politics and power than a sincere attempt to get America back on track. This disgust has emerged as a major unifying theme of the Tea Party, an incipient, grass roots political movement that exploded upon the national scene in 2009. By September 12, 2009, the Tea Party movement had grown so large that over a million tea partiers invaded D.C., marching through the halls of Congress waving signs declaring "What Would Jefferson Do?"; "I'm Taking Back My Country, one politician at a time"; "More Government for The People = Less Freedom of The People"; "Read the Tenth Amendment"; and "Wake Up America, Before Your Liberty is Gone."[11] Tea Party enthusiasm drove the results of the 2010 mid-term elections, when Tea Party issues and candidates propelled the Republican Party to recapture control of the U.S. House of Representatives and gain six U.S. Senate seats.[12]

Tea Party opposition to bailouts, stimulus packages and health-care reform is reflected in various proposals to amend the Constitution, including proposals to require a balanced budget, repeal the Sixteenth and Seventeenth Amendments, and give states a veto power over federal laws (the so-called Repeal Amendment).[13] Liberals have decried the Tea Party's

---

9. *See* Patient Protection and Affordable Care Act, Pub. L. No. 111-148, 124 Stat. 119; CNN Political Unit, *CNN Poll: Time doesn't change views on health care law*, POLITICALTICKER... (Mar. 23, 2011, 5:30 AM), http://politicalticker.blogs.cnn.com/2011/03/23/cnn-poll-time-doesnt-change-views-on-health-care-law/.

10. *See* DEMOCRATIC POLICY COMM., *The Individual Responsibility Policy Is Constitutional* (Oct. 15, 2010), http://dpc.senate.gov/dpcdoc.cfm?doc_name=fs-111-2-163.

11. *See* Lisa Miller, *223 Tea Party Signs and Placard Ideas*, TEA PARTY WDC (June 13, 2009, 1:35 PM), http://www.meridianteaparty.com/tea-party-rally-sign-ideas.

12. *See generally* Michael Cooper, *Victories Suggest Wider Appeal of Tea Party*, N.Y. TIMES, Nov. 3, 2010, at 1; *The U.S. Constitution as a Celebrity: The Rising Star of the Tea Party Isn't a Person, It's a Document*, THE TORONTO STAR, Nov. 13, 2010, at IN1.

13. *See generally* Elizabeth Wydra & David Gans, CONSTITUTIONAL ACCOUNTABILITY CTR., *Setting The Record Straight: The Tea Party and The Constitutional Powers of the Federal Government*, (July 16, 2010), http://www.theusconstitution.org/upload/fck/file/File_storage/Setting%20the%20Record%20Straight%20Issue%20Brief%20formatted(1).pdf?phpMyAdmin=TzXZ9IzqiNgbGqj5tqLH06F5Bx; Randy Barnett, *The Case for the Repeal Amendment*, 78 TENN. L. REV. 815 (2011).

call for constitutional amendments as hypocritical. They do not understand how Tea Partiers can simultaneously pledge fealty to the Constitution and seek to change it. In the words of one recent liberal blogger, this is akin to "wrapping themselves in the rhetoric of the Constitution while simultaneously trying to remake [the] document into something completely unrecognizable."[14]

These criticisms have rhetorical appeal but no real substance. Imagine that I have a bicycle that I hold dear. You borrow it one day and, not revering it as much as I, damage its seat, handlebars and spokes. When I now look at my beloved bicycle, I am deeply saddened by these changes. What should I do: lament the harm done to these important features, or restore the bicycle to its original glory? Of course I should restore it, and this is precisely what federalism-based proposed amendments seek to do for the Constitution. Restoring the Constitution, not remaking it, is the goal of the Tea Party and federalism-based amendments.

II. AMENDMENTS VERSUS REVOLUTIONS

It is worth briefly pondering the range of options for effectuating constitutional change. On one end of the spectrum is revolution, a self-conscious rebellion against an existing legal regime. The signers of the Declaration of Independence recognized and invoked the natural right of revolution in their quest to break free from the tyranny of King George III:

> We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty and the pursuit of Happiness. That to secure these rights, Governments are instituted among Men, deriving their just powers from the consent of the governed,—That whenever any Form of Government becomes destructive of these ends, it is the Right of the People to alter or to abolish it, and to institute new Government, laying its foundation on such principles and organizing its powers in such form, as to them shall seem most likely to effect their Safety and Happiness.[15]

Because revolutions are fundamental or radical breaks with existing legal regimes, they are often accompanied by violence, as was the case with the American Revolution. There is no inherent necessity for violence in the context of revolution. Theoretically, any overt, self-conscious rejection of the binding authority of an existing legal regime would qualify for the revolution label. Yet, precisely because revolutions are so complete in their rejection of existing legal authority, they are much rarer than discrete acts—

---

14. Ian Millhiser, *Cantor Endorses Bizarre Tea Party Constitutional Amendment*, THE WONK ROOM (Dec. 1, 2010, 7:00 PM), http://wonkroom.thinkprogress.org/2010/12/01/repeal-cantor.

15. THE DECLARATION OF INDEPENDENCE para. 2 (U.S. 1776).

such as constitutional amendments—that express disagreement only with distinct aspects of an existing regime. Of course not all legal regimes provide a mechanism for amendment, but most do, including our own. When discrete options are available, they provide an important pressure valve for effectuating changes broadly supported by the citizenry. But amendments are not always enough, and revolution is sometimes inevitable. In the words of the signers of the Declaration of Independence:

> Prudence, indeed, will dictate that Governments long established should not be changed for light and transient causes; and accordingly all experience hath shewn, that mankind are more disposed to suffer, while evils are sufferable, than to right themselves by abolishing the forms to which they are accustomed. But when a long train of abuses and usurpations, pursuing invariably the same Object evinces a design to reduce them under absolute Despotism, it is their right, it is their duty, to throw off such Government, and to provide new Guards for their future security.[16]

The message seems to be this: Longstanding governments should be tinkered with when desired and discarded *in toto* only when necessary to defend individuals' natural rights to life, liberty, and the pursuit of happiness. When government becomes destructive of such rights, however, revolution is not only morally just, but morally imperative.

On what side of the revolution-amendment line do we find America today? This question is harder to answer than it initially seems. Everyone agrees that the Declaration of Independence was an act of revolution. But what about a constitutional convention, like the 1787 convention in Philadelphia? Is an Article V constitutional convention an act of revolution, or a mere act of amendment? It is an interesting question because current calls for constitutional change advocate not only discrete constitutional amendments—e.g., balanced budgets or a presidential line item veto—but also the use of constitutional conventions to ratify such amendments.[17]

Because there is a noticeable states' rights undercurrent to recent calls for constitutional reform, proponents are not content with the usual proposal by two-thirds of Congress followed by ratification by three-quarters of the states.[18] Instead, they seek to bypass Congress to the extent permitted by Article V, invoking a mode of amendment never actually used. Specifically, Article V recognizes that constitutional amendments can be proposed not only upon approval by two-thirds of both houses of Congress, but also "on the application of the legislatures of two thirds of the several States."[19] It further declares that, upon application from the requisite

---

16. THE DECLARATION OF INDEPENDENCE para. 2 (U.S. 1776).

17. *See generally* David B. Rivkin, Jr. & Lee A. Casey, *The States Can Check Washington's Power*, WALL ST. J., Dec. 22, 2009, at A23.

18. *Id.*

19. U.S. CONST. art. V.

two-thirds of states, Congress "shall call a Convention for proposing Amendments" which are then deemed ratified upon approval by three-fourths of the states.[20]

But why bypass the usual process of congressional proposal and instead call for a state-initiated constitutional convention? David Rivkin and Lee Casey, two of the earliest proponents of this method, asserted in a December 2009 *Wall Street Journal* op-ed that a state-initiated convention is needed.[21] They contend that an unchecked expansion of federal power since the mid-1800s has created a situation in which Congress "has little interest in proposing limits on its own power."[22] As Bradford Plumer recently confessed in *The New Republic*, "It's difficult to imagine [two-thirds] of senators signing up to . . . suddenly make their jobs contingent on the whims of a bunch of state legislators by axing the Seventeenth Amendment."[23]

Plumer's observation is undoubtedly correct. After ratification of the Bill of Rights in 1791 and the Eleventh Amendment in 1798, every single constitutional amendment proposed by Congress affecting the vertical distribution of power—other than the Twenty-first Amendment's repeal of Prohibition—has either restricted state power or enlarged federal power.[24] Three-quarters of the states ratified each of these power-adjusting amendments, which is understandable given the specific contexts in which they were considered. The Civil War Amendments, for example, were designed to address specific state abuses of the natural and civil rights of citizens. The cumulative effect of these amendments, particularly when combined with the impact of the Supreme Court case law vastly expanding federal power, has created the impression that something radical needs to be done to rebalance sovereignty. Invoking Article V's procedure for a state-initiated constitutional convention provides a mechanism for venting these concerns.

Rivkin and Casey did not stop with advocating a state-initiated constitutional convention.[25] They proposed that states demand that the convention take up a specific constitutional amendment "to permit two-thirds of the states to propose amendments directly."[26] Amending Article V in this way, if successful, would provide a more powerful means to jumpstart consideration of states' rights by bypassing Congress completely. Rivkin and Casey confessed that their proposal is designed to "shift the power calculus" back toward the states, "enabl[ing] the states to check

---

20. *Id.*

21. *See* Rivkin & Casey, *supra* note 17.

22. *Id.*

23. Bradford Plumer, *The Revisionaries: The Tea Party's Goofy Fetish for Amending the Constitution*, THE NEW REPUBLIC, Sept. 23, 2010, at 16.

24. *See generally id.*

25. *See* Rivkin & Casey, *supra* note 17.

26. *Id.*

Washington power [and] provide a constructive outlet for much of the growing anger—especially evident in phenomena such as the 'tea party' movement—toward the political elites of both parties."[27]

The use and potential expansion of state-initiated constitutional conventions has evoked visceral opposition, mostly from the political left.[28] Dylan Matthews, writing in the *Washington Post* in April 2010, asserted that "a convention would just bring trouble" because the "flavor of the week culture war amendments—school prayer, same-sex marriage, flag-burning . . . could sneak out of a constitutional convention."[29] Even more boldly, Matthews admitted that "[t]he Constitution could use serious reform, but the institutional changes of the type procedural-minded liberals advocate don't have the constituency that silly and reckless proposals do."[30]

Putting aside Matthews's obvious political bias and his concomitant fear about "silly and reckless" proposals that do not reflect his views, Matthews's concerns about a "runaway" convention are widely shared by individuals across the political spectrum.[31] By their very nature, constitutional conventions have the potential to run away. The thirteen original states agreed to send delegates to the Philadelphia Convention with, in the words of the Continental Congress resolution calling the convention, the avowed purpose of "revising the Articles of Confederation, and reporting to congress and the several legislatures, such alterations and provisions therein, as shall . . . render the federal Constitution, adequate to the exigencies of government, and the preservation of the Union."[32] What emerged was not exactly a revision of the Articles of Confederation; so if history is any indication, it would be impossible to force a convention to focus only on specific proposals.

We could also expect special interest groups to descend on a modern convention like flies on a carcass. But it could not be held behind closed doors the way the 1787 Convention was, so every move a modern convention makes would be tweeted, blogged, and Facebooked to death, not to mention commented on by every cable pundit left and right. This alone would dampen tendencies toward adopting truly wild proposals. Moreover, the convention's proposals would still need to be ratified by three-quarters of the states. As James LeMunyon has noted, "there are a sufficient number of 'red' and 'blue' states to block any attempt to amend the Constitution in a radical way from the left or right."[33]

---

27. *Id.*

28. *See, e.g.*, Dylan Matthews, *Would a Constitutional Convention Get It Right?*, EZRA KLEIN (Apr. 1, 2010, 8:36 AM), http://voices.washingtonpost.com/ezra-klein/2010/04/would_a_constitutional_convent.html.

29. *Id.*

30. *Id.*

31. *See generally id.*

32. THE FEDERALIST NO. 40 (James Madison) (George F. Hopkins ed., 1802).

33. James M. LeMunyon, *A Constitutional Convention Can Rein In Washington*,

Pragmatically, LeMunyon's observation is probably correct. In theory, however, a constitutional convention could change the Article V rules requiring ratification of three-fourths of the states.[34] Indeed, Article VII of the U.S. Constitution declared that ratification of nine out of the thirteen states would be sufficient, even though Article XIII of the Articles of Confederation required unanimous state approval for any amendments.[35]

The 1787 Convention's blatant disregard of Article XIII, combined with its failure to stick to "revising" the Articles of Confederation, could be perceived as placing the Convention on the revolutionary side of the line. These revolutionary possibilities are admittedly possible with any convention. Yet it is also worth noting that Article XIII of the Articles of Confederation was only about ten years old when it was ignored by the Philadelphia Convention.[36] None of the provisions of the Articles of Confederation had sufficient time to become deeply rooted, venerable constitutional doctrine. Article V, by contrast—including its requirement of ratification by three-fourths of the states—is venerated, and accordingly something a modern constitutional convention would be highly unlikely to disregard.

The bottom line is that there is growing interest in using the state-initiated constitutional convention process to implement amendments rebalancing the vertical division of power. While a constitutional convention via Article V is "scary" because it has never been officially used, it offers a creative solution for vetting growing popular concern about ever-expanding centralized power.

III. PROPOSALS TO REBALANCE SOVEREIGNTY

Aside from proposals to invoke a state-initiated constitutional convention, current proponents of constitutional change have advocated specific amendments geared towards rebalancing sovereignty between the state and federal governments. I will discuss only two of the most provocative proposals, both of which come from Professor Randy Barnett: (1) a Federalism Amendment; and (2) a Repeal Amendment.

A. The Federalism Amendment

Professor Barnett's Federalism Amendment contains a cornucopia of items, the bulk of which can be characterized as attempts to restore federalism, including repealing the Sixteenth Amendment, limiting the exercises of the spending power to those necessary to carry out enumerated powers, and trimming back the commerce power by forbidding Congress

---

WALL ST. J., Apr. 1, 2010, at A19.

34. *See* U.S. CONST. art. V.

35. *See* U.S. CONST. art. VII; ARTICLES OF CONFEDERATION art. XIII.

36. *See generally* ARTICLES OF CONFEDERATION art. XIII.

from regulating activity "wholly within a single state, regardless of its effects outside the state or whether it employs instrumentalities therefrom."[37] Critics of these vertical separation provisions have pointed out that repealing the income tax power might have undesirable economic consequences and that forbidding regulation of activity "wholly within a single state" does not forbid much at all.[38] Of course, cutting back on federal spending power would inevitably cause massive withdrawal symptoms from addicted states.

Despite these criticisms, one has to give Barnett credit for moving the federalism ball down the field. Rather than simply complaining about things, he has generated extensive discussion about why rebalancing sovereignty is needed, and how to go about it. At the same time, Barnett's Federalism Amendment includes some provisions that, at least at first glance, seem to have nothing to do with restoring federalism. For example, section one would grant Congress expanded power over any interstate or foreign activity that is not technically "commerce," which Barnett explains was designed to give Congress power to regulate activities such as pollution.[39] The implications of section one are potentially quite far-reaching. Other than pollution, Barnett offers no other clarification regarding what sort of new activities would fall under this expanded congressional power, obscuring its original public meaning.[40]

The Federalism Amendment would also expand federal judicial power to include "the power to nullify any prohibition or unreasonable regulation of a rightful exercise of liberty" and mandates that the Constitution be "interpreted according to [its] public meaning" at the time of the relevant text's enactment.[41] These last two provisions contained in section five are clearly designed to carry out Barnett's articulated vision of a "presumption of liberty" and its relationship to original public meaning. While they are not necessarily "federalist" in nature, section five's proposals are critically important to a robust understanding of the nature of sovereignty. As I have advocated elsewhere, the Framers did not merely divide the sovereignty pie among the states and federal government; they also reserved a good deal of sovereignty to "We the People."[42] This is the message of the Ninth Amendment, which declares that the enumeration of certain rights—e.g., the Bill of Rights—"shall not be construed to deny or disparage others

37. Randy E. Barnett, *The Case for a Federalism Amendment*, WALL ST. J., Apr. 23, 2009, at A17.

38. *See, e.g.*, Ilya Somin, *Randy Barnett's 'Federalism Amendment,'* THE VOLOKH CONSPIRACY (Apr. 23, 2009, 3:08 PM), http://volokh.com/posts/1240513704.shtml.

39. *See id.*

40. *See* Barnett, *supra* note 37.

41. *Id.*

42. ELIZABETH PRICE FOLEY, LIBERTY FOR ALL: RECLAIMING INDIVIDUAL PRIVACY IN A NEW ERA OF PUBLIC MORALITY 10–15 (2006).

retained by the people."[43] It is also one of the messages in the Tenth Amendment, which states that any power not given to the federal government is "reserved to the States respectively, or to the people."[44] Taken together, the Ninth and Tenth Amendments stand for the proposition that the people—the original repository of all sovereignty—retained all rights and power not specifically ceded to state or federal government.

Therefore, in understanding federalism, it is critically important not to forget the role of the people. Yes, the sphere of state sovereignty is significant and must be protected against encroachment by the federal government, but section five of Barnett's Federalism Amendment serves as an important reminder that "We the People" should not become lost in a quest to restore the vertical balance of power. Properly conceived, rebalancing sovereignty is not just a matter of redrawing lines of power between the federal and state governments. It is also a matter of triangulating—i.e., making sure that individual rights and power, as well as state sovereignty, are respected.

B. The Repeal Amendment

Another amendment that has gained a good deal of traction—more so than the larger and more complex Federalism Amendment—is the so-called Repeal Amendment. In September 2010, Randy Barnett and William Howell, Speaker of the Virginia House of Delegates, wrote an op-ed in the *Wall Street Journal* advocating a brief amendment as follows:

> Any provision of law or regulation of the United States may be repealed by the several states and such repeal shall be effective when the legislatures of two-thirds of the several states approve resolutions for this purpose that particularly describe the same provision or provisions of law or regulation to be repealed.[45]

As Barnett and Howell explained, the purpose of the Repeal Amendment is to give a supermajority of states the power to veto federal legislation and regulations that are widely unpopular—health care reform obviously comes to mind—providing a "new political check" on a "runaway federal government."[46] They acknowledge that even if states vetoed a federal law, Congress would be free to reenact it with a simple majority.[47] Even so, the Repeal Amendment would serve an important deterrent effect, requiring Congress to consider states' reactions to legislation before passage and, if

---

43. U.S. CONST. amend. IX.

44. U.S. CONST. amend. X.

45. Randy E. Barnett & William J. Howell, *The Case for a 'Repeal Amendment,'* WALL ST. J., Sept. 16, 2010, at A23; *see also* Barnett, *supra* note 13.

46. *Id.*

47. *See id.*

ultimately repealed by the states, "forc[ing] Congress to take a second look at a controversial law."[48]

Critics initially worried that the Repeal Amendment's reference to "[a]ny provision of law or regulation of the United States" would be sufficiently broad to permit states to veto treaties or even provisions of the Constitution with the support of only two-thirds of states, effectively bypassing Article V.[49] However, Barnett has subsequently acknowledged that the Repeal Amendment's reference to "law[s] . . . of the United States," was intentionally borrowed from the Supremacy Clause,[50] which makes three clear textual distinctions between the Constitution, "laws of the United States" and treaties made under authority of the United States.[51] As such, the Repeal Amendment would be limited to repealing federal statutes and agency regulations.

Critics have also lambasted the Repeal Amendment based on the fact that it gives equal weight to the opinions of small and large states. For example, Professor Sanford Levinson has called the Repeal Amendment a "'really terrible idea'" because it would give "'outsize influence'" to "'small parochial rural states in which most Americans do not live.'"[52] The fact that Levinson does not like giving equal weight to small and large states tells us that he's not a big fan of the concept of federalism. Moreover, his use of the pejorative adjective "parochial" to describe small states reveals a common liberal bias against rural America, which liberals fault for clinging too tightly to guns, Bibles, and the Constitution. It is much better, under this elitist liberal view, to let densely populated, "sophisticated" urban areas dominate the legal system.

The Repeal Amendment has been introduced thus far by legislators in twelve states who are planning to use the state-initiated constitutional convention process to force its consideration.[53] It has also been introduced in the Senate by Republican Senator Mike Enzi of Wyoming and the U.S. House of Representatives by Republican Congressman Bob Bishop of Utah.[54] House Majority Leader Eric Cantor of Virginia has praised it warmly, calling it a way to "provide a check on the ever-expanding federal

---

48. *Id.*

49. *See, e.g.*, Bob Marshall, *Why Meddle With the Constitution?*, RICHMOND TIMES-DISPATCH, Oct. 17, 2010, *available at* http://www2.timesdispatch.com/news/2010/oct/17/ed-marshall17-ar-566354.

50. Randy Barnett, *The Case for a 'Repeal Amendment,'* THE VOLOKH CONSPIRACY, (Sept. 16, 2010, 6:00 AM), http://volokh.com/2010/09/16/the-case-for-a-repeal-amendment.

51. U.S. CONST. art. VI.

52. Kate Zernike, *Proposed Amendment Would Enable States to Repeal Federal Law*, N.Y. TIMES, Dec. 20, 2010, at 14.

53. *See id.*

54. S.J. Res. 12, 112th Cong. (2011); H.R.J. Res. 62, 112th Cong. (2011). Additionally, Rep. Bishop introduced it in the 111th Congress. H.R.J. Res. 102, 111th Cong. (2010).

government, protect against Congressional overreach, and get the government working for the people again, not the other way around."[55] Though it will be difficult to garner the support of two-thirds of the states to call a constitutional convention—and it would even harder to garner the support of two-thirds of both houses of Congress—there seems to be enough growing support for the Repeal Amendment that its overarching message about the need to rebalance sovereignty will somehow find a way to be meaningfully manifested. Even something as simple as requiring every federal bill to cite a specific constitutional power source—a promise made in the Republicans' recent Pledge to America—could help.[56]

IV. CONCLUSION

We live in fascinating times. Rarely, if ever before, have so many Americans talked so much about the Constitution. Never in my lifetime did I think I would witness popular media and grassroots, non-lawyer political activists discussing and debating the Commerce Clause, the Necessary and Proper Clause, the taxing and spending power, and the need for restoring federalism. One of my neighbors recently sheepishly pulled out of his coat a pocket Constitution, smudged with fingerprints and underlined in places. He wanted to talk about Obamacare and the constitutional bases for lawsuits challenging it. This is the gift the Tea Party movement has given us: It has made it acceptable and fashionable again to talk about the Constitution. While some elitists may whine that these pesky Americans do not know what they are talking about and should not have any input, my own experience is that they know more than many lawyers do, and are hungry to learn more. This cannot, by any stretch of the imagination, be a bad thing for America.

The reason why federalism-based constitutional amendments are being widely proposed, discussed and debated is because it does not take a degree in rocket science (or its rough equivalent, law) to realize that the federal government's powers have spun out of control. Supreme Court interpretations of some of the most important constitutional provisions defining the division of power between people, states and federal government have cumulatively eroded the fundamental architecture of the Constitution itself. The Federalism Amendment, the Repeal Amendment, the balanced budget amendment and others are designed to restore this architecture, rebalancing sovereignty in the name of protecting "We the

---

55. Evan McMorris-Santoro, *Cantor Urges "Open Mind" on VA Legislature Plan to Blow Up the Constitution*, TPMDC (Nov. 30, 2010, 1:45 PM), http://tpmdc.talkingpointsmemo.com/2010/11/va-legislator-says-gop-congress-wants-to-help-him-deconstruct-the-consitution-video.php.

56. *See* David Weigel, *Republicans Start Teaching Members How to Obey the Constitution*, SLATE (Dec. 20, 2010, 1:49 PM),http://www.slate.com/blogs/blogs/weigel/archive/2010/12/20/republicans-start-teaching-members-how-to-obey-the-constitution.aspx.

People." Before you drink the mainstream media Kool-Aid and dismiss these efforts as right wing, anachronistic, or just plain silly, ask yourself whether you want more or less liberty. If you want more (and I suspect you do), remember that our Constitution created a federal, not national, government for this very reason. Aside from the hopeless cynics among us, liberty is never silly.

REOPENING THE CONSTITUTIONAL ROAD TO REFORM: TOWARD A SAFEGUARDED ARTICLE V CONVENTION

MICHAEL STERN*

"[A] constitutional road to the decision of the people, ought to be marked out, and kept open, for certain great and extraordinary occasions."
–James Madison, *The Federalist No. 49*[1]

Every one of the twenty-seven amendments to the United States Constitution has been proposed by the Congress.[2] Even though the First Congress proposed a number of amendments that limited congressional powers or privileges (namely the Bill of Rights[3] and the amendment to limit congressional pay raises[4]), subsequent Congresses have shown little interest in following this example. They have proposed amendments that significantly expand congressional power (such as the Sixteenth Amendment that authorized a federal income tax[5]) but have proposed none that significantly limit congressional power or prerogatives. Recent Congresses, for example, have declined to propose amendments to require a balanced budget or impose term limits.[6] This would have come as no surprise to the Framers, who understood that Congress could not be expected to provide a check on itself.[7] The system they designed not only divided powers within the federal government, but also between the federal and state governments to provide a "double security" for the rights of the people.[8] As James Madison explained in *The Federalist No. 51*, under this

---

* The author formerly served as Senior Counsel in the Office of General Counsel, U.S. House of Representatives, as Deputy Staff Director for Investigations for the Senate Committee on Homeland Security and Governmental Affairs, and as Special Counsel to the House Permanent Select Committee on Intelligence. He writes about constitutional issues relating to the legislative process and other congressional legal issues at www.pointoforder.com.

1. THE FEDERALIST NO. 49, at 108 (James Madison) (J. & A. McLean ed., 1788).

2. *See* Paul G. Kauper, *The Alternative Amendment Process: Some Observations*, 66 MICH. L. REV. 903, 904 (1968).

3. U.S. CONST. amends. I–X.

4. U.S. CONST. amend. XXVII.

5. U.S. CONST. amend. XVI.

6. *See* Michael B. Rappaport, *Reforming Article V: The Problems Created by the National Convention Amendment Method and How to Fix Them*, 96 VA. L. REV. 1509, 1513 (2010).

7. *See id.* at 1525.

8. *See* THE FEDERALIST NO. 51, at 119–20 (James Madison) (J. & A. McLean ed., 1788).

system "[t]he different governments will control each other."[9] For this reason they included in Article V of the Constitution an alternative method for proposing constitutional amendments, one that did not require congressional acquiescence.[10] The convention method of amendment gave the states a constitutional road to bypass Congress when it was necessary to "erect barriers against the encroachments of the national authority," as Alexander Hamilton wrote in *The Federalist No. 85*.[11]

However, uncertainties and fears regarding the convention method have prevented its successful use to propose constitutional amendments.[12] In particular, many have feared that an Article V Convention might stray far from the concerns that caused the states to call for it.[13] The states might desire to set forth on the road to a specific constitutional reform, but a so-called "runaway convention," it is suggested, could take an unforeseen and dangerous detour from the intended path, proposing radical or ill-considered amendments to the Constitution.[14]

In this Article, I will evaluate the risks of a runaway convention in light of the constitutional text, structure, and purpose of Article V and will suggest why these risks are much smaller than often suggested. I will also suggest additional safeguards to minimize any concerns regarding a runaway convention. In combination with the inherent protections of Article V, such safeguards can ensure that the constitutional road to reform will be clearly defined and well marked, and may be traveled safely by the states when they must act to impose limitations on a "runaway Congress."

I. CONSTITUTIONAL CONSIDERATIONS

Article V provides that:

> [T]he Congress, whenever two thirds of both Houses shall deem it necessary, shall propose Amendments to this Constitution, or, on the Application of the Legislatures of two thirds of the several States, shall call a Convention for proposing Amendments, which, in either Case, shall be valid to all Intents and Purposes, as Part of this Constitution, when ratified by the Legislatures of three fourths of the several States, or by Conventions in three fourths thereof, as the one or the other Mode of Ratification may be proposed by the Congress; Provided that no

---

9. *Id.* at 120.

10. *See* Rappaport, *supra* note 6, at 1516–17.

11. THE FEDERALIST NO. 85, at 363–64 (Alexander Hamilton) (J. & A. McLean ed., 1788).

12. *See, e.g.*, Michael Stokes Paulsen, *A General Theory of Article V: The Constitutional Lessons of the Twenty-Seventh Amendment*, 103 YALE L.J. 677, 763 (1993).

13. *See generally, e.g.*, Arthur H. Taylor, *Fear of an Article V Convention*, 20 BYU J. PUB. L. 407 (2006) (analyzing the rationality of common fears related to the process).

14. *See, e.g.*, Gerald Gunther, *The Convention Method of Amending the United States Constitution*, 14 GA. L. REV. 1, 4–5 (1979).

> Amendment which may be made prior to the Year One thousand eight hundred and eight shall in any Manner affect the first and fourth Clauses in the Ninth Section of the first Article; and that no State, without its Consent, shall be deprived of its equal Suffrage in the Senate.[15]

The debate involving the risk of a runaway convention has generally focused on the question of whether a "Convention for proposing Amendments" is, by its constitutional nature, an unlimited convention or whether such a convention may be limited, as a matter of constitutional theory, to considering only such amendments within the scope of the "Application" of the states. Some commentators suggest that unless one can provide a definitive answer to this legal question, it is simply too risky to hold an Article V Convention.[16] I maintain that this is not the case. Nonetheless, the constitutional foundations of the Article V Convention are significant insofar as they shed light on how the constitutional actors in the convention amendment process should, and likely will, fulfill their roles.

A. *The Origins of the Article V Convention*

Article V originated as part of the Virginia Plan presented to the Philadelphia Convention on May 29, 1787.[17] The Virginia Plan stated that the "Articles of Union" should be amendable "whensoever it shall seem necessary, and that the assent of the National Legislature ought not be required thereto."[18]

This provision was referred to the Committee of Detail, which produced a draft stating that "[t]his Constitution ought to be amended whenever such Amendment shall become necessary; and on the Application of the Legislatures of two thirds of the States in the Union, the Legislature of the United States shall call a Convention for that Purpose."[19] Implicit in this statement is that state legislatures would determine, at least in the first instance, when it would become necessary to amend the Constitution and that a convention would be called for the purpose of considering any amendment that the states deemed necessary.[20]

---

15. U.S. CONST. art. V.

16. *See, e.g.*, Gunther, *supra* note 14, at 25 (warning that the road "promises controversy and confusion and confrontation at every turn"); Richard W. Hemstad, *Constitutional Amendment by Convention – a Risky Business*, 36 WASH. ST. B. NEWS 16, 21 (1982) (predicting the possibility of "[A] period of significant instability in the American political system . . . .").

17. *See* Douglas G. Voegler, *Amending the Constitution by the Article V Convention Method*, 55 N.D. L. REV. 355, 360–61 (1979).

18. 1 THE RECORDS OF THE FEDERAL CONVENTION OF 1787 22 (Max Farrand ed., 1911).

19. 2 THE RECORDS OF THE FEDERAL CONVENTION OF 1787 159 (Max Farrand ed., 1911).

20. The language chosen by the Committee of Detail may have been derived from the

Subsequently, on September 10, 1787, objections targeted this provision on the grounds that it gave only the state legislatures the power to initiate amendments.[21] Hamilton argued that the states would "not apply for alterations but with a view to increase their own powers."[22] Congress, he contended, "will be the first to perceive and will be most sensible to the necessity of amendments, and ought also be empowered" to call a convention on its own initiative.[23]

Madison then proposed a substitute that addressed Hamilton's concerns.[24] His proposal provided:

> The Legislature of the U— S— whenever two thirds of both Houses shall deem necessary, or on the application of two thirds of the Legislatures of the several States, shall propose amendments to this Constitution, which shall be valid to all intents and purposes as part thereof, when the same shall have been ratified by three fourths at least of the Legislatures of the several States, or by Conventions in three fourths thereof, as one or the other mode of ratification may be proposed by the Legislature of the U.S.[25]

The convention adopted the proposal by a vote of nine in favor, one opposed, and one divided.[26]

The Madison Substitute served two functions. First, it eliminated the convention altogether, reflecting Madison's reservations regarding the effectiveness of the convention method.[27] Second, it put the state legislatures and Congress on equal footing. Congress shall propose amendments whenever amendments are deemed necessary by two-thirds of both Houses or applied for by two-thirds of the states.

The Madison Substitute does not explicitly state *what* amendments Congress shall propose. The only reasonable interpretation, however, is that

---

Georgia Constitution of 1777, which stated that "the assembly shall order a convention to be called for that purpose." GA. CONST. of 1777, art. LXIII. The Georgia assembly was to call a convention for amendments "specifying the alterations to be made, according to the petitions preferred to the assembly." *Id.*; *see* RUSSELL L. CAPLAN, CONSTITUTIONAL BRINKMANSHIP: AMENDING THE CONSTITUTION BY NATIONAL CONVENTION 95 (1988).

21. *See* 2 THE RECORDS OF THE FEDERAL CONVENTION OF 1787, *supra* note 19, at 557–58.

22. *Id.* at 558.

23. *Id.*

24. *See id.* at 559.

25. *Id.*

26. *See id.*

27. Responding to the draft produced by the Committee of Detail, "Mr. Madison remarked on the vagueness of the terms, 'call a Convention for that purpose,'" posing the following questions: "How was a Convention to be formed? [B]y what rule decide[d]? [W]hat the force of its acts?" *Id.* at 558. After the convention method was reintroduced, Madison again noted "difficulties might arise as to the form, the quorum [etc.]" *Id.* at 630.

Congress is to propose those amendments deemed necessary by two-thirds of both Houses or applied for by two-thirds of the states. It would be far-fetched to contend, as literally permitted by the language, that Congress could propose an amendment that was different from one deemed necessary by two-thirds of both Houses. It would be equally unreasonable to conclude that Congress could propose an amendment that was different from one applied for by the state legislatures.[28]

There is, or at least there was at the time, a significant logistical difference between the two types of amendments. While it would have been straightforward to determine which amendments might be deemed necessary by two-thirds of Congress, coordination among the state legislatures was much more difficult considering the limitations of communications in the eighteenth century. It does not appear from the records of the Philadelphia Convention that anyone considered the possibility that the state legislatures could agree, in advance, on the text of a particular desired amendment to the Constitution. One can only assume that the Framers believed that agreement on a single text without a meeting among the states was impractical or created too great a potential for miscommunication and misunderstanding.

This view likely underlay the objection raised by George Mason, on September 15, 1787, to the Madison Substitute.[29] Mason described the provision as "exceptionable [and] dangerous" because "the proposing of amendments is in both the modes to depend, in the first immediately, and in the second, ultimately, on Congress."[30] Therefore, Mason believed that "no amendments of the proper kind would ever be obtained by the people, if the Government should become oppressive, as he verily believed would be the case."[31]

---

28. Such a reading would mean that if the states applied for an amendment to establish freedom of speech, for example, the Congress could propose, by a majority vote, an amendment on an entirely different subject, something that it would lack the power to do in the absence of the state applications. Clearly this was not the intent of the Madison Substitute. As James Kenneth Rogers has noted, the Madison Substitute makes little sense except in the context of a specific type of amendment desired by the states. Note, *The Other Way to Amend the Constitution: The Article V Constitutional Convention Amendment Process*, 30 HARV. J.L. & PUB. POL'Y 1005, 1017 (2007).

29. *See* 2 THE RECORDS OF THE FEDERAL CONVENTION OF 1787, *supra* note 19, at 629.

30. *Id.*

31. *Id.* Mason's view would be echoed in the remarks of a delegate to the state constitutional convention of Maryland two centuries later; Royce Hanson, during the debates of Maryland Constitutional Convention of 1967–68, noted that:

> [T]here is probably no group of people in creation less likely to reform themselves than the members of the legislature when the time for that reform has arrived, and it is for this reason that it seems to me that we should provide in the constitution a means external to the legislature for the revision of that part of the constitution which pertains to the legislature.

To remedy this problem, "[Gouverneur] Morris [and Eldridge] Gerry moved to amend [Madison's language] so as to require a Convention on [the] application of [two thirds] of the [states.]"[32] Madison responded that he "did not see why Congress would not be as much bound to propose amendments applied for by two thirds of the States as to call . . . a Convention on the like application."[33]

Although not reflected in the records of the Philadelphia Convention, the answer to Madison's point must have been that the calling of a convention was merely a ministerial act, with no degree of discretion, while proposing amendments would necessarily have involved some degree of discretion. For example, even if two-thirds of the states applied for a convention and clearly specified the type of amendment they wanted, Congress would still have to agree on the precise wording of the amendment. If Congress was unable to do so, the amendment would never be proposed.

Despite believing the Morris/Gerry proposal to be unnecessary, Madison stated that he had no objection to "a Convention for the purpose of amendments," although he reiterated his concerns about the effectiveness of the convention method, given that there was no definition of how the convention would actually operate.[34] Lacking time or inclination to address these concerns, the Philadelphia Convention agreed to the Morris/Gerry proposal.[35] The amendment assumed its final form when it was agreed to include substantive limitations on the amendment power, including "that no State, without its Consent, [could] be deprived of . . . equal Suffrage in the Senate."[36]

It seems evident from this history that the primary, if not sole, purpose of the convention method was to enable the states to initiate the amendment process without the need of congressional assistance and to solve the logistical problem of reaching agreement on a single text.[37] The history also suggests an intent that the Article V Convention serves as an aid to the states and not to function as an independent entity exercising significant discretion in its own right.

This view of Article V, moreover, was the one presented to the states during the ratification process. Madison continued to adhere to the view that the proposing power given to the convention was merely a quasi-

---

JOHN J. DINAN, THE AMERICAN STATE CONSTITUTIONAL TRADITION 61 (2009).

32. 2 THE RECORDS OF THE FEDERAL CONVENTION OF 1787, *supra* note 19, at 629.

33. *Id.* at 629–30.

34. *Id.*

35. *See id.*

36. *Id.* at 662–63.

37. *See* CAPLAN, *supra* note 20, at 29 ("The division of the amendment power was the essential compromise of [A]rticle V, for determining who could propose amendments went far to determining what kind of amendments would be adopted.").

ministerial extension of the state's power to initiate amendments.[38] In *The Federalist No. 43*, he explained that Article V "equally enables the general and the State governments to originate the amendment of errors."[39] In other words, there was no substantive difference between the power of the states to apply for a convention and the power of Congress to propose amendments.

During the debates over ratification of the Constitution, Federalists pointed to the convention method as a key safeguard to protect the states and the rights of the people against potential overreach by the new national government. For example, in *The Federalist No. 85*, Hamilton emphasized the convention method as a means of correcting any perceived errors in the Constitution, explaining that "alterations [in the Constitution] may at any time be effected by" the requisite number of states.[40] He explained that "whenever nine, or rather ten States, were united in the desire of a particular amendment, that amendment must infallibly take place."[41] Rejecting the notion that Congress could block the convention method, Hamilton wrote:

> [T]he national rulers, whenever nine states concur, will have no option upon the subject. By the fifth article of the plan, [C]ongress will be *obliged*, "on the application of the legislatures of two-thirds of the states, (which at present amounts to nine) to call a convention for proposing amendments, which *shall be valid* to all intents and purposes, as part of the [C]onstitution, when ratified by the legislatures of three-fourths of the states, or by conventions in three-fourths thereof." The words of this article are preemptory. The [C]ongress "*shall* call a convention." Nothing in this particular is left to the discretion of that body. And of consequence all the declamation about the disinclination to a change vanishes in air. . . . We may safely rely on the disposition of the state legislatures to erect barriers against the encroachments of the national authority.[42]

These assurances regarding the convention method would, at best, be misleading if the states lacked any ability to define or control the Article V Convention. If the proposing power of the convention were entirely separate from and independent of the application power of the states, one could not "safely rely on the disposition of the State legislatures to erect barriers against the encroachments of the national authority,"[43] nor could one say that the state and federal governments had equal ability to "originate the amendment of errors."[44]

---

38. THE FEDERALIST NO. 43, at 65 (James Madison) (J. & A. McLean ed., 1788).
39. *Id.*
40. THE FEDERALIST NO. 85, *supra* note 11, at 361.
41. *Id.* at 362.
42. *Id.* at 363–64.
43. *Id.*
44. THE FEDERALIST NO. 43, *supra* note 38, at 65.

B. Textual Analysis of Article V

The key language of Article V is that "[t]he Congress, whenever two thirds of both Houses shall deem it necessary, shall propose Amendments to this Constitution, or, on the Application of the Legislatures of two thirds of the several States, shall call a Convention for proposing Amendments . . . ."[45]

Professor Michael Stokes Paulsen, echoing Professor Charles Black, argues that "[t]he most straightforward reading of the constitutional text concerning what the convention *is*—'a Convention for proposing Amendments'—strongly suggests that it must be, in the words of Professor Black, 'a convention for proposing such amendments as that convention decides to propose.'"[46] Professor Paulsen further contends that "[t]he text supplies no basis for inferring a power, on the part of either Congress or applying state legislatures, alone or in concert, to limit what the convention may consider."[47]

It is true that the text is silent as to what amendments the convention may propose. It is not at all obvious, however, that this silence means that the convention is unlimited in what it may propose. To the contrary, it seems perfectly logical to infer a relationship between the "Application" of the state legislatures and the "Convention for proposing Amendments" to which the application gives rise.[48] Rather than reading the "Convention for proposing Amendments" as a "[c]onvention for proposing such amendments as that convention decides to propose,"[49] it would be at least equally natural to read it as a "convention for proposing such amendments as the state legislatures have applied for."[50]

Professor Paulsen also suggests that the structure of Article V supports the inference that a convention must be unlimited. In his words, "[t]he convention-proposal method is worded in parallel with the congressional-proposal method, implying an equivalence of their proposing powers . . . ."[51] Because Congress is not subject to any limitation on the amendments it may propose, in Professor Paulsen's view, the convention must be similarly unlimited.[52]

This analysis overlooks the presence of the two triggering clauses in Article V.[53] In the case of the congressional-proposal method, the triggering

---

45. U.S. CONST. art. V.

46. Paulsen, *supra* note 12, at 738 (quoting Charles L. Black, Jr., *Amending the Constitution: A Letter to a Congressman*, 82 YALE L.J. 189, 199 (1972)).

47. *Id.*

48. U.S. CONST. art V.

49. Black, Jr., *supra* note 46, at 199.

50. *Id.*

51. Paulsen, *supra* note 12, at 739.

52. *See id.*

53. U.S. CONST. art. V.

clause is "whenever two thirds of both Houses shall deem it necessary."[54] In the case of the convention-proposal method, the triggering clause is "on the Application of the Legislatures of two thirds of the several States."[55] The structure of Article V implies an equivalence between these two triggering clauses, which becomes clearer when one considers the original language of the Madison Substitute.[56] In that provision, the two triggering clauses were alternative means of triggering the congressional-proposal method.[57] As finally adopted in Article V, one clause triggers the congressional-proposal method, while the other triggers the convention-proposal method.[58]

When one recognizes the equivalence of the two triggering clauses, the structure of Article V strongly supports the conclusion that a convention may be limited.[59] Just as Congress's power to propose amendments is limited to those amendments that two-thirds of both Houses deem necessary, the convention's power to propose amendments must be limited to those amendments that two-thirds of the state legislatures have applied for.

Finally, Professor Paulsen argues that the Framers must have understood the term "convention" to refer to a body with unlimited or "plenary" powers.[60] This contention is unpersuasive for several reasons. First, the historical evidence of practice at the time of the founding generation suggests that conventions served a variety of purposes and the term did not have a single fixed meaning.[61] Specifically, not all conventions were understood to be plenary, and limited conventions were known—such as the convention provided for in the Georgia Constitution of 1777.[62]

---

54. *Id.*

55. *Id.*

56. *See* 2 THE RECORDS OF THE FEDERAL CONVENTION OF 1787, *supra* note 19, at 559.

57. *See id.*

58. *See* U.S. CONST. art. V.

59. *See id.*

60. Paulsen, *supra* note 12, at 740 ("[T]he best early evidence of 'contemporaneous understanding,' as revealed by early practice, suggests that the founding generation understood conventions to be plenary.").

61. *See* CAPLAN, *supra* note 20, at 3–26. Indeed, Madison's objection to "the vagueness of the terms, 'call a Convention for the purpose'" strongly suggests that the meaning of the term in the context of Article V was not so clear or self-evident. 2 THE RECORDS OF THE FEDERAL CONVENTION OF 1787, *supra* note 19, at 558.

62. *See* CAPLAN, *supra* note 20, at 95–98. Recent scholarship by Professor Robert Natelson further supports this point. *See* Robert. G. Natelson, *Amending the Constitution by Convention: A Complete View of the Founders' Plan (Part 1 in a 3 Part Series)*, POLICY REPORT NO. 241 (GOLDWATER INSTITUTE), Sept. 2010, at 8–12, *available at* http://www.goldwaterinstitute.org/article/5005. Surveying the historical evidence, Professor Natelson concludes that "[a] reference to a 'convention' in an 18th-century document did not necessarily mean a convention with plenary powers, even if the reference was in a constitution. Although it *might* refer to an assembly with plenary powers, it was *more likely* to denote one for a limited purpose." *Id.* at 10.

Second, even if conventions generally had been understood to be plenary, it does not follow that the specific "Convention for proposing Amendments" established in Article V was intended to be of this nature.[63] This convention, after all, was intended for a specific, and limited purpose—to propose amendments to "this Constitution."[64] It was not given the power to enact anything, merely to propose, and the power to propose was limited to "amendments" to "this Constitution."[65] Even the power to propose was subject to substantive limits.[66] For example, it could not extend to denying the states equal suffrage in the Senate.[67] The evidence, therefore, does not support the conclusion that an Article V Convention must be understood as plenary.

C. The Purpose of the Article V Convention

Scholars who believe that an Article V Convention must be unlimited have struggled to explain the constitutional purpose that would be advanced by this interpretation. Although it is possible to argue that the unlimited convention is simply an unintended consequence of the compromise language that the Framers ultimately settled upon, this argument is weakened by the absence of a plausible rationale for the unlimited convention.[68]

This issue must be distinguished from questions regarding the practical difficulties of defining and enforcing limits on an Article V Convention. It is one thing to argue that these difficulties mean that an Article V

---

63. *See generally* Gunther, *supra* note 14.

64. U.S. CONST. art. V.

65. I will not rehearse here the long-standing debate as to whether the Philadelphia Convention itself was a "runaway convention" that ignored the limits on its authority under the Articles of Confederation. Fears that an Article V Convention might exercise power beyond that granted by Article V itself are, by definition, extra-constitutional in nature. No one can prove definitively that a group of individuals will not claim to exercise some authority that they do not have. It should be observed, however, that the chances of an Article V Convention having the prestige or ability to assert an extra-constitutional legitimacy, in effect to proclaim a new constitutional order for the United States, is exceedingly remote.

66. U.S. CONST. art. V

67. *Id.*

68. As Professor Rappaport notes:

> If limited conventions are not recognized by the Constitution, then the constitutional provision allowing the states to decide whether to hold a convention seems peculiar. Why would the Constitution allow the states to decide to call a convention, but not allow them to specify what subjects the convention should discuss?

Rappaport, *supra* note 6, at 1521.

Convention will be unlimited as a practical matter. It is another to contend that Article V affirmatively grants a convention the power to address any subject, however unrelated to the application that gave rise to that convention. Other than to discourage state legislatures from applying for a convention in the first place, it is difficult to see what purpose is served by granting the convention such wide powers of proposal.

It might be argued that the Framers chose an unlimited convention because, on the one hand, they saw little risk in allowing the convention to propose whatever amendments it pleased, while, on the other, attempting to define the limits of an Article V Convention in any kind of useful way would simply be too difficult. This argument has some attraction, particularly if one believes, as I do, that the ratification requirements of Article V constitute substantial protection against radical or ill-conceived amendments.

There are, however, two strong objections to this argument. First, the Framers were not as blithe toward proposed constitutional amendments as it would suggest. Article V requires a two-thirds majority of both Houses to propose a constitutional amendment, even though the amendment must still be ratified by three-fourths of the states.[69] It is difficult to see why the Framers would not have insisted that an amendment proposed by a convention be similarly grounded in a broad consensus—as would be the case if the amendment were responsive to the application of two-thirds of the state legislatures.

Second, the difficulty of definition may explain why Article V does not attempt to define the relationship between the state application and amendments proposed by convention for purposes of *all* conventions that might be applied for by the states. It does not, however, provide a reason why constitutional actors[70] in the amendment process could not define and enforce such a relationship in the context of a *particular* convention call.

Other attempts to identify a constitutional purpose of the unlimited convention are similarly unavailing. Professor Walter Dellinger argues that "the [F]ramers did not want to permit enactment of amendments by a process of state proposal followed by state ratifications without the substantive involvement of a national forum."[71] By transferring the proposing power from Congress to the convention, the Framers chose a body that would be "like Congress, a deliberative body with a national perspective, capable of assessing the need for constitutional change as well as developing proposals to be submitted for ratification."[72]

It is possible that the Framers valued the deliberative capabilities of the convention, although there is no evidence of this in the debates during the

---

69. *Id.*

70. State legislatures, the courts, Congress, and the convention itself.

71. Walter Dellinger, *The Recurring Question of the "Limited" Constitutional Convention*, 88 YALE L.J. 1623, 1630 (1979).

72. *Id.* at 1626.

Philadelphia Convention or the ratification process. To the contrary, the evidence discussed above suggests that the purpose of the deliberation process was to serve primarily as an aid to the states in solving the logistical difficulties of reaching an agreement on the text of a proposed amendment.[73]

In the event that the convention was to exercise a significant deliberative function, it does not follow that its deliberations should be unlimited. It is possible that the Framers intended the convention to deliberate on alternative solutions to pertinent issues; however, it is difficult to imagine what purpose would be served by having the convention deliberate on unrelated issues. Not only would such a broad deliberative scope serve no discernible purpose, it would make it less likely that the convention would fulfill what Professor Dellinger acknowledges as its core mission of responding to the states' grievances.[74]

Like Professor Dellinger, Professor Gerald Gunther emphasizes the deliberative function of the Article V Convention, but he also suggests that the convention serves the purpose of providing a check on the less deliberative proceedings of the state legislatures.[75] He notes that "[t]hirty-four state legislatures acting separately simply are not as likely to act as seriously as a single national forum in the proposing of constitutional amendments."[76] Professor Gunther contends that this consideration supports the interpretation of the convention as unlimited.

There is little evidence to suggest that the Article V Convention was intended to provide a check on the state legislatures. Professor Gunther cites Roger Sherman's objection, raised after the Philadelphia Convention had adopted the Madison Substitute, "that three fourths of the States might be brought to do things fatal to particular States."[77] Contrary to Gunther's assertion, Sherman's objection was not to the Madison Substitute in particular, as shown by the fact that he continued to raise objections after

---

73. Indeed, Professor Dellinger acknowledges that the amendment-proposing function does not necessarily involve any significant degree of deliberation. He notes that the "most plausible reading" of the Madison Substitute "is that it would have permitted two-thirds of the state legislatures to propose amendments to the Constitution; Congress would merely transmit those amendments to be ratified." *Id.* at 1628. Moreover, he acknowledges that the transfer of the amendment-proposing function from Congress to the convention "may have been based on Mason's belief in the practical necessity of having a single deliberative body undertake the consultation, debate, drafting, compromise, and revision necessary to produce an amendment." *Id.* at 1629–30.

74. *See id.* at 1639 ("It is reasonable to expect that a convention would choose to confine itself to considering amendments addressing the problem that led states to apply for the convention.").

75. *See* Gunther, *supra* note 14, at 12–13.

76. *Id.* at 19.

77. *Id.* at 15 (internal quotation marks omitted).

the convention method was adopted.[78] What Sherman wanted was substantive limits on the amendment power to protect states' rights.[79]

Granting Professor Gunther's premise that the Framers intended the convention as a check on the states, the rationale for an unlimited convention is still lacking. Even a convention that is limited to consideration of a single amendment must deliberate regarding the meaning and effect of that amendment and reach a decision as to whether to propose it.[80] Thus, assuming for argument's sake that the Framers intended that the Article V Convention serve as a check on the allegedly impulsive state legislatures, it fulfills that purpose just as well within the framework of a limited convention as that of an unlimited convention.

Finally, it has been argued that the unlimited convention is a necessary result of the Framers' desire to limit Congress's role in the convention method process.[81] Professor Paulsen, for example, argues that "[i]f states could call for a limited convention, Congress would be placed in the position of prescribing and enforcing . . . limitations on the work of the convention, giving Congress a major role inconsistent with the convention method's intended purpose."[82]

The convention method was designed to limit Congress's role in the state-initiated amendment process.[83] Allowing Congress to define the limits of an Article V Convention would indeed raise serious concerns. However, no such concerns are raised if the states prescribe the limits in their application and Congress simply calls the convention, without adding to or subtracting from what the states have declared. In fact, were Congress to reject the application for a limited convention, or call for an unlimited convention in contravention of the application, this would, itself, arguably expand Congress's role beyond what the Framers intended.[84]

With regard to determining whether a proposed amendment must be submitted to the states for ratification, Congress will have to exercise some degree of judgment, regardless of whether a convention is limited or unlimited. There could, for example, be disputes about whether a particular amendment was proposed in accordance with the convention's voting or other rules. Similarly, Congress may have to resolve disputes about whether a particular amendment falls within the scope of a limited convention. Such a determination, however, need not involve an undue amount of

---

78. 2 THE RECORDS OF THE FEDERAL CONVENTION OF 1787, *supra* note 19, at 630–31.

79. *See id.*

80. *See* Dellinger, *supra* note 71, at 1631–32.

81. *See* Paulsen, *supra* note 12, at 739.

82. *Id.* at 739.

83. *See* Paulsen, *supra* note 12, at 739.

84. To be clear, if one assumes that an application for a limited convention is invalid, Congress presumably would have the power to reject such application. But the fact that Congress is required to determine whether an application is valid is not an argument for or against a limited convention. *See* Dellinger, *supra* note 71, at 1624.

congressional discretion. If the states set forth clear rules defining the scope of the convention, Congress may enforce these rules without raising any concerns about exceeding its proper role.[85]

D. The Role of Constitutional Doubt

The above discussion identifies some weaknesses of the theory that an Article V Convention must be unlimited and explains why the limited convention theory is more consistent with the constitutional text, structure, and purpose. It must be acknowledged, however, that the *purely legal* issue of whether an Article V Convention may be limited cannot be definitively resolved. Constitutional scholars have long debated the question, and it is widely recognized to be a quintessentially open one.

Our concern here, however, is not with identifying the "right answer" to a constitutional question in the abstract, but with determining the real-world risks of a runaway convention. Those who are worried about a runaway convention will probably not be mollified by the assurance that such a convention would be unconstitutional, even if there were greater scholarly consensus on the point. Moreover, asking the question of how the United States Supreme Court might resolve the issue produces no more of a definitive answer, and indeed, it is unclear when or whether the courts might intervene in the convention amendment process.[86]

It has often been assumed that these uncertainties enhance the risks of an Article V Convention, but this assumption is flawed. What it overlooks is the role of constitutional doubt in guiding the actions of constitutional actors, other than the courts, within the framework of the convention amendment process. These actors must exercise both political and legal judgment in performing their functions. So long as there is a serious doubt regarding the constitutionality of an out-of-scope amendment, the

---

85. *Cf.* United States v. Rostenkowski, 59 F.3d 1291, 1306 (D.C. Cir. 1995) (holding that a court may interpret and apply a rule of the U.S. House of Representatives without infringing on the House's exclusive rulemaking power, so long as the rule is sufficiently clear that the court may be confident in its interpretation).

86. As Professor Randy Barnett has observed, claiming that something is "unconstitutional" usually means one of the following: (1) it may refer to the actual meaning of the Constitution, independent of any authority's interpretation of that meaning; (2) it may refer to what the Supreme Court has said about a particular constitutional issue in the past; or (3) it may refer to a prediction that a majority of the Supreme Court would vote that the particular action is unconstitutional. *See* Randy Barnett, *In What Sense is the Personal Health Care Mandate "Unconstitutional"?*, THE VOLOKH CONSPIRACY (Apr. 16, 2010, 11:27 AM), http://volokh.com/2010/04/16/in-what-sense-is-the-personal-health-insurance-mandate-unconstitutional. In this case, however, there is virtually no relevant judicial authority and little basis for predicting how, or whether, the Supreme Court would rule. We are therefore primarily interested in the best arguments as to the meaning of the Constitution and how constitutional actors, other than the courts, will likely respond to them.

constitutional actors should refrain from proposing, submitting, or ratifying such an amendment.

1. The Article V Convention. If the state application limits the convention's deliberations either to a particular subject or a particular amendment, the convention will have to determine how to respond to that limitation. The issue is likely to arise at the outset of the convention, when the delegates vote to adopt rules to govern the proceedings. As discussed later, the states applying for a limited convention should instruct their delegates to vote for rules that limit the convention's deliberations in accordance with the application.

As a practical matter, the question of the constitutionality of an out-of-scope amendment will probably be of limited significance to the Article V Convention as a whole. Lacking any extended institutional existence, it is doubtful that the convention would give a great deal of attention to the constitutional issue, unless there was a serious attempt to push an out-of-scope amendment. In that case, it seems likely that the political difficulties of proposing the amendment would have greater salience than the legal issues.

Those delegates who have been instructed to comply with the limitations set forth in the application of their state, however, will have a strong legal incentive to abide by those instructions. Failure to do so would mean violating a personal obligation under state law. Unless the delegate believes that the United States Constitution clearly overrides this obligation, the delegate would likely comply with it. Furthermore, it should be noted that even if the Article V Convention had the power, under the federal Constitution, to propose out-of-scope amendments, it does not follow that states are powerless to instruct their delegates with regard to such amendments.[87] Thus, the legal uncertainties weigh heavily against any delegates violating their state law obligations to oppose an out-of-scope amendment.

2. Congress. If an Article V Convention were to propose an out-of-scope amendment, Congress would have to decide whether to submit the amendment to the states for ratification. Such submission cannot occur automatically because, under Article V, Congress must determine whether ratification will take place by state conventions or legislatures—as has been the case for all congressionally proposed amendments except for the Twenty-first Amendment.

Members of Congress take an oath to support the Constitution and are generally thought to have a duty not to vote for unconstitutional measures.[88]

---

87. Professor Paulsen, for example, notes that the applying states, in his view without power to limit the convention directly, "might well exercise considerable control by selecting delegates committed to enforcing a limitation on the agenda." Paulsen, *supra* note 12, at 760.

88. *See Oath of Office*, UNITED STATES SENATE, http://senate.gov/artandhistory/history/common/briefing/Oath_Office.htm#1 (last visited Apr. 5, 2011).

Although the nature of this obligation and the quality of Congress's compliance with it have been the subject of considerable debate, it is likely that most members of Congress would feel themselves obligated to ensure that only valid amendments are submitted to the states for ratification. Furthermore, Congress has an institutional incentive to limit the authority of an Article V Convention with respect to proposing amendments. Finding that an Article V Convention could not be limited would give that convention a greater authority to propose amendments than Congress itself, since the latter can only propose amendments when two-thirds of both Houses deem it necessary.

Congress also has an incentive to act in advance of actually receiving an out-of-scope amendment. By declaring that it will not submit out-of-scope amendments for ratification, Congress would both deter any such amendments and avoid subsequent charges that its refusal to submit a particular amendment was based on policy preference, rather than constitutional principle.

It seems unlikely that many members of Congress would favor, as a matter of policy, an unlimited Article V Convention. Nevertheless, some members may believe that the Constitution requires that an Article V Convention be so unlimited. Alternatively, those members could support a constitutional amendment recently introduced in Congress that would remove any doubt that an Article V Convention may be limited to consideration of a single constitutional amendment.[89]

3. The States. If Congress were to submit an out-of-scope amendment for ratification by state legislatures, state legislators would face the same constitutional issue as members of Congress.[90] State legislators also take an oath to uphold the Constitution of the United States. State legislators who voted to apply for an Article V Convention limited to a single subject or amendment would arguably violate this oath if they subsequently voted to ratify an out-of-scope amendment.[91]

State legislatures have a substantial interest in avoiding this situation because ratifying an out-of-scope amendment might undermine future attempts to call a limited Article V Convention. Accordingly, as discussed later, state legislatures may adopt procedures that would make it virtually impossible to ratify out-of-scope amendments. This pre-commitment can ensure that subsequent political pressure to ratify a popular out-of-scope

---

89. *See* H.R.J. Res. 95, 111th Cong. (2010) (known as the "Madison Amendment").

90. It is theoretically possible, but highly unlikely, that Congress could submit an out-of-scope amendment for ratification by state conventions. As discussed later, the state legislatures can erect legal barriers to protect against this remote possibility.

91. The state legislator's duty to reject an out-of-scope amendment does not necessarily turn on whether the legislator voted for a limited Article V Convention in the first place. However, it would be difficult for a legislator to reconcile a vote for a limited convention with a subsequent vote to ratify an amendment that exceeded the scope of that limited convention.

amendment will not undermine the constitutional position of state legislatures.

State legislatures are in a different position than Congress in one respect. While Congress has a constitutional duty to submit a valid proposed amendment for ratification, the state legislatures are under no such duty to ratify such an amendment. Thus, constitutional doubt as to the validity of an out-of-scope amendment cuts only one way—against ratification.

II. EVALUATING THE RISK OF A "RUNAWAY CONVENTION"

At this point, we should define more precisely what is meant by a "runaway convention." At the extreme, the phrase implies a convention that adopts radical or far-reaching proposals, such as repealing the Bill of Rights or similar outlandish measures. Those who suggest such a possibility warn that the absence of legal certainty regarding the outer scope of a convention's power means that there is no such thing as a "safe" Article V Convention.

The question must be asked: "safe compared to what?" After all, somewhere in our constitutional system must lie the ultimate authority to make law and declare what the law is. This power, wherever it resides, necessarily implies the possibility of results that we would regard as unacceptable.

Judicial review, for example, creates the risk that the Constitution will effectively be changed or "amended" whenever a majority of the Supreme Court decides that it should be.[92] Whether one views any particular decision of the Court as unjustified or unacceptable, it is impossible to deny that judicial review creates the risk of extreme or unacceptable outcomes.

On the other hand, limiting or eliminating judicial review, while reducing the risk of "judicial amendments" to the Constitution, would increase the risk that the political branches would violate or ignore constitutional limits on their authority. Professor John Hart Ely paraphrases the critics of his theory of judicial review thus: "[Y]ou'd limit courts to the correction of failures of representation and wouldn't let them second-guess the substantive merits? Why, that means you'd have to uphold a law that provided for _____!"[93] In other words, minimizing the risk of a runaway court means, to some extent, increasing the risk of a runaway legislature.

Assessing the risk of a runaway convention must therefore include consideration of not only the risks that may exist in using the convention method of amendment, but also the risks that might be reduced by the

92. *See* Taylor, *supra* note 13, at 415 ("[O]ur system already includes a wide-open amendment proposing process through the judiciary.").

93. JOHN HART ELY, DEMOCRACY AND DISTRUST A THEORY OF JUDICIAL REVIEW 181 (1980).

method's use or by the mere recognition of the method as usable. These offsetting risks are, of course, precisely those for which the Framers designed the Article V Convention in the first place. It can scarcely be denied that the limited powers granted to the Congress in Article I of the Constitution have not proved to be a meaningful check on the expansion of federal power. The Article V Convention, if available as intended to check the "encroachments of the national authority," would mitigate this risk.

Of course, if one does not believe that the growth of federal power is a matter of concern, then one may not wish to take any risks, however minimal, to counteract it.[94] In that case, however, the real objection is to the existence of the convention method of amendment. Fear of a runaway convention, while reducing the risk that an Article V Convention will be called or even creditably threatened, in the short term, does not change the fact that the convention method of amendment is unquestionably a part of the Constitution. Insisting on the unlimited nature of the Article V Convention also increases the risk, whatever it may be, that someday such an unlimited convention will occur.

A. The Inherent Safeguards of Article V

Because no convention has ever been called under Article V and the process for selecting delegates is as yet undefined, it is relatively easy to stoke fears that the convention might fall under the control of radical or irresponsible elements prone to the temptation of a runaway convention. Yet sober reflection reveals that this danger is more imagined than real.

Although some state legislatures might choose a different method, it is likely that most delegates to an Article V Convention will be elected by popular vote.[95] Political scientists Paul J. Weber and Barbara A. Perry argue that the process of selecting delegates to an Article V Convention can be predicted with a reasonable degree of confidence.[96] Candidates for election "will include those who have an active interest in the purpose of the convention and who are willing to take a position for or against

---

94. *See* Jack M. Balkin, *The Consequences of a Second Constitutional Convention*, BALKINIZATION (Sept. 17, 2010, 4:49 PM), http://balkin.blogspot.com/2010/09/consequences-of-second-constitutional.html (noting that whether one thinks an Article V Convention "is a good thing or a bad thing has much to do with whether you think that the convention will address and help resolve serious issues that the country needs to face down").

95. The great weight of opinion in modern times has favored election of convention delegates. *See, e.g.*, Sam J. Ervin, Jr., *Proposed Legislation to Implement the Convention Method of Amending the Constitution*, 66 MICH. L. REV. 875, 892 (1968) (noting that legislation introduced by Senator Ervin to govern Article V Convention proceedings initially allowed either election or appointment of delegates but was changed to require election). Delegates to the majority of state constitutional conventions have also been popularly elected. *See* DINAN, *supra* note 31, at 12.

96. *See* PAUL J. WEBER & BARBARA A. PERRY, UNFOUNDED FEARS: THE MYTHS AND REALITIES OF A CONSTITUTIONAL CONVENTION 113–15 (1989).

amendments."[97] They are likely to have substantial name recognition, organizational and financial support, and prior campaign experience.[98] In the course of campaigning, they will be asked to take positions on proposed amendments and whether they would take part in a runaway convention.[99] Those elected will generally reflect mainstream political views, be representative of existing political interests, and will be "highly unlikely to approve radical changes."[100]

Therefore, even apart from outside constraints on an Article V Convention, the chances of delegates approving outlandish types of amendments are highly remote. But it must be remembered that an Article V Convention has only the power to *propose* amendments. It cannot actually affect any change to the Constitution without the subsequent ratification of three-fourths of the states. Thus, the inherent safeguards in the Article V process include:

> [T]he number of delegates and divisions within the convention itself, which would make it extraordinarily difficult for one faction or a radical position to prevail; the delegates' awareness that the convention results must be presented to Congress, which might not forward any amendment that went beyond the convention mandate; the Supreme Court, which might well declare certain actions beyond the constitutional powers of the convention; and *most important of all, the need to get the proposed amendment ratified not only by the thirty-four states that called for the convention, but by thirty-eight states.*[101]

Noting that "[m]ore effective constraints on a constitutional convention can hardly be imagined,"[102] Weber and Perry conclude that, "[n]otwithstanding the arguments of legal scholars with limited methodological tools (or partisan objectives) and political columnists with active imaginations, calling a constitutional convention would be a safe political process."[103] Before his appointment to the bench, Justice Antonin Scalia similarly observed that the risk of an "open convention" is "not much of a risk" since "[t]hree-quarters of the states would have to ratify whatever came out of the convention."[104]

The safeguards inherent in the Article V Convention process apply to all potential amendments, but they particularly ensure that a convention will

---

97. *Id.* at 113.
98. *See id.*
99. *See id.* at 114.
100. *Id.* at 115.
101. *Id.* at 117 (emphasis added).
102. *Id.*
103. *Id.* at 119–20.
104. Antonin Scalia, *Supplement at the American Enterprise Institute Forum, in* A CONSTITUTIONAL CONVENTION: HOW WELL WOULD IT WORK? 22–23 (Am. Enter. Inst. for Pub. Policy Research, 1979), *quoted in* CAPLAN, *supra* note 20, at 138.

not adopt radical, divisive, or controversial proposals.[105] It might be argued, however, that an Article V Convention could still propose out-of-scope amendments of a different type. For example, a convention might make hasty or ill-considered changes to the text of an amendment contained in the state applications, with unintended consequences. Or, a convention might be faced with a temporary groundswell of support for a particular amendment, say, for instance, in reaction to an unpopular Supreme Court decision, causing it to exceed the mandate set forth by the applying states. These more realistic possibilities may necessitate that additional safeguards be built into the process.

B. Additional Safeguards

To build additional safeguards into the Article V Convention process, the states applying for the convention must agree on and set forth in their applications the text of the single amendment they wish the convention to consider. Without such a text, a convention nominally limited to a particular topic is unlikely to be, in practice, significantly more limited than an unlimited convention. Judging whether a proposed amendment falls within a particular topic is ultimately a subjective exercise that is vulnerable to manipulation or obfuscation. Just as the enumerated powers of the Congress under Article I have proved to be a weak barrier against expansion of the federal government, so might a convention limited to a single subject, such as a "balanced budget," expand into unforeseen areas.[106]

It should be noted here that some commentators believe that, although the Article V Convention may be limited to a particular subject or topic, it cannot be limited solely to considering a specific amendment.[107] The distinction appears to be based on the idea that limiting the convention to a single amendment unduly restricts its deliberative freedom and effectively transfers the proposing power from the convention to the states.[108]

My own view is that this distinction, while attractive on the surface, is neither ultimately persuasive nor particularly workable. First of all, limiting

---

105. Even Professor Gunther, who warns against the risks of an Article V Convention, acknowledges that it is unlikely to adopt "wild-eyed proposals." Gunther, *supra* note 14, at 10.

106. *See id.* at 18 ("If a convention cannot be limited to simply voting 'yes' or 'no' on a particular balanced budget scheme, what is to prevent it from considering such questions as permissible or impermissible expenditures for, say, abortions or health insurance or nuclear power?"). This is not to say that a limited convention would necessarily expand in such a way, but the primary constraints would be the inherent safeguards of Article V rather than any additional legal or procedural safeguards created by specifying a particular subject matter.

107. *See, e.g.*, Ervin, *supra* note 95, at 884.

108. *See id.*

the convention to a single text does not prevent it from fully deliberating about the particular amendment. It is not clear why deliberating about a single text is any less deliberative than deliberating about a more broadly defined subject. Since the convention retains the ultimate decision as to whether to propose the amendment, a single amendment rule also does not transfer the proposing power to the states.

Second, limiting the convention to a single amendment is simply a way of narrowly defining the subject that the convention shall consider. If the state application for a convention defines the "subject" by reference to the text of a specific amendment, it is difficult to see how this categorically changes the nature of the convention. A rule giving the convention the deliberative freedom to consider alternative solutions to a particular problem would lead to endless debate whether the "problem" was defined so narrowly as to deprive the convention of the appropriate amount of deliberative freedom.

No constitutional principle appears to support distinguishing a convention limited to a single subject from one limited to a single amendment. The only justification for rejecting the narrower limitation would seem to be one of efficiency—if the convention rejects the particular amendment on the grounds that there is a superior solution, the states would have to submit a new application to permit consideration of the alternative. Efficiency, however, clearly was not the objective of Article V. Moreover, nothing in Article V requires the states to limit the convention to a particular amendment—it simply permits them to do so.

Accordingly, I concur with the view of Professor William Van Alstyne that an Article V Convention limited to the text of a single amendment is perfectly permissible.[109] Moreover, having the states submit such an amendment in their application would seem to address the criticism of the convention-method process that the states are too cavalier in applying for conventions.[110] If the states do the hard work of hammering out and agreeing on the text of a single amendment, they are far more likely to take the process seriously and use it only advisedly.

Nevertheless, the fact that the states propose a single amendment does not necessarily mean that the convention must be without any power to change it. The state legislatures could provide a channel by which minor and non-controversial changes could be adopted—for example, by unanimous consent of the convention—and thereby minimize constitutional objections without significantly increasing the risk of a runaway convention.

In order to ensure that an Article V Convention is limited to consideration of a single amendment identified by the states in their applications to Congress, the states may employ the following safeguards.

---

109. *See* William W. Van Alstyne, *Does Article V Restrict the States to Calling Unlimited Conventions Only?—A Letter to a Colleague*, 1978 DUKE L.J. 1295, 1305 (1978).

110. *See* Gunther, *supra* note 14, at 3.

These safeguards could be embodied in a uniform act similar to other uniform acts created to enable the states to exercise their federal functions.[111]

The Application Safeguard. In applying for an Article V Convention, each state legislature applying may pass a resolution containing the identically worded text of the amendment sought. The applications should specify (1) that they are to be considered only in conjunction with other applications seeking the identical amendment and (2) that the convention shall be for the sole purpose of considering the specified amendment.

The applications may also provide Congress with a period of time—for example, six months from the date on which the required thirty-four applications have been received—in which to propose an identical constitutional amendment pursuant to Article V's congressional method. If Congress acts, the applications will be voided and no convention will be required.

The Convention Safeguard. Each applying state will require its delegates to vote for convention rules that limit its deliberations to consideration of the single amendment at issue. As noted previously, such rules may permit looking beyond the stipulated amendment only if the convention complies with rigorous procedural requirements, such as for a unanimous vote of the convention.[112] These rules, adopted at the convention's outset, may also provide that the convention proceedings will terminate after an up-or-down vote on the amendment.

The Delegate Safeguard. Each state may require its delegates to support the specified rules and limit their participation in the convention to consideration of the specified amendment. Violation of this pledge might be made punishable by sanctions, disqualification, or both.

The Congressional Safeguard. Although Congress's role in the convention process is largely ministerial, Congress remains responsible for submitting any proposed constitutional amendments to the states for ratification and for determining the method of ratification. The applying states may request that Congress refuse to submit any out-of-scope amendment for ratification.

This safeguard would be further enhanced if Congress pre-committed not to submit an out-of-scope amendment for ratification. Congress could take this action either by joint resolution or by a resolution adopted by the House, the Senate, or both. Even a commitment by a single House would offer substantial assurance that an out-of-scope amendment would not be submitted for ratification. The resolution could be adopted with respect to a

---

111. *See, e.g.*, UNIFORM FAITHFUL PRESIDENTIAL ELECTORS ACT (Interim Draft Mar. 2, 2010), *available at* http://www.jamesmadisoncenter.org/PresidentialElectors/NCCUSLPropo sedFaithlPresElectors.pdf.

112. Where an amendment is changed in accordance with such a procedural requirement, the modified amendment would continue to be considered an "in scope" amendment for purposes of subsequent ratification.

specific convention call. Alternatively, the states could adopt a uniform act establishing the procedures for the Article V Convention application, thereby enabling Congress to adopt a resolution regarding any "out-of-scope" amendment as defined by that uniform act.

The Ratification Safeguard. The most important safeguard, of course, is the one specifically provided by the Framers, namely that no amendment proposed by the convention is valid until ratified by three-fourths (thirty-eight) of the states. Needless to say, it is exceedingly unlikely that any applying state would ratify an out-of-scope amendment.

To further assure applying states that their sister states will not ratify an out-of-scope amendment, each applying state might adopt measures to prevent such an eventuality. State legislatures could adopt rules requiring a supermajority to ratify an out-of-scope amendment or stipulating that consideration of such an amendment is entirely out of order.[113] More controversially, a legislature might prohibit any state convention for the purpose of ratifying an out-of-scope amendment.[114]

The Judicial Safeguard. As a last resort, an out-of-scope amendment could be challenged in federal court. Such a challenge would, of course, raise significant justiciability issues, but enabling legislation could remove all non-constitutional barriers to such a suit. Thus, while there is no guarantee that the courts would reach the merits, proponents of an out-of-scope amendment would face a substantial risk that their efforts would be struck down by the courts.

III. CONCLUSION

The full power of the above safeguards is evident in their cumulative impact, as illustrated by the difficult road faced by a proponent of an out-of-scope amendment. In order to obtain the convention's endorsement of such an amendment, its proponent must first persuade a majority of the convention to defeat the convention rules and vote in favor of the out-of-scope amendment. This would mean persuading delegations from at least ten applying states to violate their oaths and risk legal sanctions, not to mention bad publicity. In addition, costly and protracted litigation would likely ensue in the respective state courts of the ten "faithless" delegations.

Second, the proponent of an out-of scope amendment must persuade Congress to submit the amendment for ratification, in clear violation of the

---

113. One federal court has held that states have significant latitude in determining the procedures for ratifying a federal constitutional amendment. *See* Dyer v. Blair, 390 F. Supp. 1291, 1307 (N.D. Ill. 1975) (Future Supreme Court Justice John Paul Stevens authored the opinion).

114. Such a provision, which would become relevant only in the unlikely event that Congress chose the convention method of ratification, would present perhaps the most likely scenario under which federal courts might reach the merits of whether an out-of-scope amendment is constitutionally valid.

applying states' intentions and possibly in violation of Congress's own commitment not to do so.

Third, the proponent must to convince thirty-eight states to ratify the out-of-scope amendment. This would require ratification by at least twenty-two of the applying states. In order to have any prospect of accomplishing such a feat, the proponent would have to overcome state rules prohibiting ratification or establishing supermajority requirements of both houses in those twenty-two states to ratify the amendment. Alternatively, the proponent would have to believe that Congress would choose the convention method of ratification—which it has done only for ratification of the Twenty-first Amendment—and would have to have a legal strategy to require states to call such conventions.

Finally, the prospect of a federal court challenge would remain. Whatever its ultimate outcome, such a challenge would be time-consuming and expensive for the proponents of the out-of-scope amendment.

Given this outlook, it is impossible to imagine that anyone would seek to hijack a convention for purposes of promoting an out-of-scope amendment. If one hypothesizes an out-of-scope amendment so broadly popular as to have even a remote chance of surmounting the obstacles we would erect, there would be far easier ways to achieve the desired goal.

In short, these safeguards will keep the constitutional road to reform marked and open and will secure it against any chance of unwanted detours by a so-called "runaway convention."

CONVENTIONAL WISDOM: ACKNOWLEDGING UNCERTAINTY IN THE UNKNOWN

MARY MARGARET PENROSE*

I. INTRODUCTION

Necessity never made a good bargain.[1]

That which has never been tried or tested cannot be confidently, much less boastfully, touted. In fact, something that has never occurred requires careful assessment and often receives numerous differing predictions regarding success and potential failure. In employing the scientific model, one relies on a hypothesis to calculate what is most likely to occur. But law does not use the true scientific method. Rather, law is as fluid and changing as the participants who make and interpret it.[2] Thus, as the Article V issue of a State-Convention process to amend the United States Constitution is considered, all commentators on the topic must confess that the dialogue being proffered is, at best, mere forecasting.[3] In truth, as none of the twenty-seven constitutional amendments have even been proffered through the State-Convention method, even the most sage constitutional scholars are at a loss to know, with any real precision, what will occur or which

---

* Professor of Law, Texas Wesleyan School of Law. The author wishes to thank her 2009 and 2010 Constitutional Law classes for insisting that the Constitution is a fluid and "living" document. Their faith in our democracy convinces me that our Constitution will live on for generations to come. The author further wishes to thank the entire library staff at Texas Wesleyan, with particular thanks to Laura Fargo, for their unconditititional research assistance. Essays like this could not be drafted without the aid and support of librarians.

1. BENJAMIN FRANKLIN, POOR RICHARD'S ALMANACK 10 (Blackwell North America, Inc. 1987).

2. While *Marbury v. Madison* reminds that "it is emphatically the province and the duty of the judicial department to say what the law is," such declarations are far from static. Marbury v. Madison, 5 U.S. 137, 177 (1803). Society has, on occasion, witnessed radical shifts in court doctrines over the period of just a few years. *See, e.g.*, Lawrence v. Texas, 539 U.S. 558 (2003); Bowers v. Hardwick, 478 U.S. 186 (1986); West Coast Hotel v. Parrish, 300 U.S. 379 (1937); Lochner v. New York, 198 U.S. 45 (1905).

3. *See, e.g.*, Gerald Gunther, *The Convention Method of Amending the United States Constitution*, 14 GA. L. REV. 1, 2 (1979) ("[C]onstitutional convention route bristles with unanswered questions."). Professor Gunther went further to proclaim his belief that "the convention route promises uncertainty, controversy, and divisiveness at every turn." *Id.* at 5. In closing his essay, Professor Gunther later confessed: "Everything I have said constitutes conjecture about the past and advice about the future." *Id.* at 25. *See also* Neal S. Manne, *Good Intentions, New Inventions, and Article V Constitutional Conventions*, 58 TEX. L. REV. 131, 135 (1979) ("For the constitutional law scholar, the consideration of the convention alternative is a foray into conjecture and speculation.").

bodies—executive, legislative or judicial—will actually be involved in the process.[4]

Much like Benjamin Franklin's admonishment, we must recognize that going into an Article V State-Convention scenario without any guiding principles could lead to a very unstable and unpredictable outcome.[5] While this is not, in itself, problematic, legislatures and judges should heed the warning: "necessity never made a good bargain."[6]

In this short Essay, I will respond to the honorable Michael Stern's article—*Reopening the Constitutional Road to Reform: Toward a Safeguarded Article V Convention*[7]—that assures us there is nothing to fear from the State-Convention process. While this is surely one approach, I believe it to be too trusting, if not naïve, in light of our constitutional history. With literally nothing serving as our compass, we risk the creation of rules that will undoubtedly be borne out of necessity. If nothing changes to guide the process, we can only hope that those in power during such an unprecedented and momentous undertaking will make wise and limited decisions. But, as this Essay demonstrates, there is nothing mandating such behavior. The State-Convention model has never been tested or used. Therefore, no one can be certain that upon its invocation either the procedures utilized or the outcomes reached will be moderate or even moderated.

II. VISIONARY IDEAS — SHORT ON DETAIL

[O]ne cannot work in constitutional law for long without appreciating the hazards of guesses about the future.[8]

---

4. *See* Gunther, *supra* note 3, at 10. Professor Gunther observed that it remains "a real question as to whether the courts would consider this an area in which they could intervene; other aspects of the amendment process have been held by the courts to raise non-justiciable questions." *Id.* Professor Gunther's further concern is expressed as follows:

> In any event, the prospect of [any] lawsuit simply adds to the potentially divisive confrontations along the convention road—a confrontation between Court and Congress, to go with the possible other confrontations, between Congress and the convention, between Congress and the states, and perhaps between the Supreme Court and the states.

Id.

5. *Id.* at 10–11. Professor Gunther admonished that the State-Convention method, in its current from without any guiding principles "[I]s a road that promises controversy and confusion and confrontation at every turn." *Id.* at 25.

6. FRANKLIN, *supra* note 2, at 10.

7. Michael Stern, *Reopening the Constitutional Road to Reform: Toward a Safeguarded Article V Convention*, 78 TENN. L. REV. 765 (2011).

8. Gunther, *supra* note 3, at 1.

The Constitutional Convention that brought forth the United States Constitution was as visionary as it was revolutionary.[9] Historians and statesmen continue to revel in the Constitution's enduring value, its historic brilliance, and legal sustainability. Foremost among the visionary components of the Constitution is the recognition that to endure, the document must be capable of change.[10] Change was explicitly provided for in the Constitution through the amendment process in Article V.[11] However, necessarily lacking in this otherwise visionary proposal is any detailed provision for how the State-Convention option should logistically operate.[12]

Article V reads, in pertinent part, as follows:

> The Congress, whenever two thirds of both houses shall deem it necessary, shall propose Amendments to this Constitution, or, on the Application of the Legislatures of two thirds of the several States, shall call a Convention for proposing Amendments, which, in either Case, shall be valid to all Intents and Purposes, as Part of this Constitution, when ratified by the Legislatures of three fourths of the several States, or by the Conventions in three fourths thereof, as the one or the other Mode of Ratification may be proposed by the Congress . . . .[13]

And, while this language provides certain mathematical provisions that many Americans find familiar (because generally it takes two-thirds to propose and three-fourths to ratify) any guiding principles are patently lacking. Thus, the fear of a "runaway convention" appears at least as rationally based as the confidence shown by those debunking the idea of a "runaway convention." In truth, both arguments are mere prognostication and neither can be supported with traditional authority. There are few cases on these issues, and the cases that do exist are inconclusive and, at times, inconsistent. This schism exists simply because the State-Convention model has never been used, never been tested, and presents Americans with the potential, just as real as any other alternative, that a radical State-Convention paradigm could rethink the entire United States Constitution.[14]

Historically, the Framers' debate seemed to evidence concern that the amendment process not be entirely placed with either the individual states or the national government. Members of the Constitutional Convention

---

9. *See* Ralph R. Martig, *Amending the Constitution, Article Five: The Keystone of the Arch*, 35 MICH. L. REV. 1253, 1253–57 (1936).

10. *See id.* at 1253.

11. *See* U.S. CONST. art. V.

12. *See* Clifton McCleskey, *Along the Midway: Some Thoughts on Democratic Constitution Amending*, 66 MICH. L. REV. 1001, 1003 (1967).

13. U.S. CONST. art. V.

14. *See* Martig, *supra* note 9, at 1256 ("There can be no doubt that [the Constitutional] convention, by proceeding to draft a new frame of government, exceeded its powers; these were explicit and confined to the sole purpose of revising the Articles of Confederation.").

evidenced distrust for any one government being the sole repository for the Amendment process. Thus, after much crafting and compromise,[15] Article V, in its current form, was presented and ratified without much fanfare. To date, Article V has been used twenty-seven times, though the first twelve amendments were passed nearly contemporaneously with the original Constitution.[16] At other times, like in the Civil-War era, Amendments were passed in a grouping of just a few years.[17] Then, there is the Twenty-seventh Amendment, which, curiously enough, was originally offered in 1789 and deemed ratified in 1992.[18]

Furthermore, all twenty-seven Amendments except the Twenty-first Amendment were actually ratified by the State-Legislature model.[19] That

---

15. *See* Manne, *supra* note 3, at 142–46.

16. *See* Martig, *supra* note 9, at 1266; *see also* Robert G. Dixon, Jr., *Article V: The Comatose Article of our Living Constitution*, 66 MICH. L. REV. 931, 931 (1967) ("[L]ay[] aside the ten in the Bill of Rights, which were really a continuation of the original process of constitution-making . . . .").

17. *See* Dixon, *supra* note 16, at 931 ("[T]he three Civil War amendments, which were part of the unique process of reformation of the Union . . . ."); *see also* Manne, *supra* note 3, at 132–33 ("This paucity of formal change appears more acute if one considers, entirely reasonably, the first ten amendments (the Bill of Rights) to have been a continuation of the original process of constitution making. Three other amendments (the Thirteenth, Fourteenth, and Fifteenth) originated in the aftermath of the Civil War, and reform following military suppression of revolution does not fit very neatly into traditional doctrines of American constitutionalism. Two amendments—the Prohibition and repeal amendments—effectively cancel each other out.").

18. Congress presented the Twenty-seventh Amendment with twelve other amendments. *See* VARAT, ET AL., CONSTITUTIONAL LAW: CASES AND MATERIALS 16 n.1 (12th ed. 2006). In 1791, ten of the twelve amendments were ratified and comprise what is now considered the "Bill of Rights." *Id.* In 1992, the thirty-eighth state, Michigan, finally ratified the Twenty-seventh amendment, reaching the requisite three-fourths states needed for the Amendment to take effect. *Id.* There remains, however, some controversy as to whether the twenty-seventh Amendment is truly part of the Constitution. *See id.*

19. *See Amendment of the Constitution by the Convention Method Under Article V*, AMERICAN BAR ASS'N SPECIAL CONSTITUTIONAL CONVENTION STUDY COMM., at 1–2 (1973); *see also* Martig, *supra* note 9, at 1270 (commenting that one of the founders, Senator Ferry of Connecticut, cautioned that the convention method was not preferable because conventions are "dilatory, expensive, and unwise"). In an admonishment that remains timely, and apropos to the question at hand regarding State Conventions, Senator Ferry warned:

> If a convention is once assembled you cannot limit its power to the simple amendment which you are proposing to it. It may go on to amend your State constitution and to subvert the whole machinery of your State government, and there is no power in your State to stop it.

Id. at 1272. Martig also explained that "[i]n submitting the Twenty-first Amendment, Congress failed to provide any regulation for the calling and supervising of the conventions,

amendment, the lone and anomalous amendment ratified through the State-Convention mode of unfortunately offered little guidance.

There is no discernable reason that the States have never reached the required two-thirds mandated for Congress to call an Article V Convention.[20] Every state has at one time or another petitioned Congress to call an Article V Convention.[21] The states have been close to reaching the requisite number and continue, even today, to send resolutions calling for an Article V Constitutional Convention.[22] But, having come up short on every occasion, the State-Convention procedure has never been tested.

While it is likely that most, beyond legislators and constitutional scholars, are unaware of the State-Convention model, society must be prepared to address this contingency should it occur. Presuming, for present purposes, that at some point in the future the State-Convention method will be used, I must respectfully disagree with Stern in his assessment that sufficient safeguards currently exist. I also disagree that the primary controversy is whether a convention will be limited to considering only discretely proposed amendments.[23]

Instead, I perceive the controversy as being much broader in scope and the challenges ahead much more deeply imbedded in our constitutional fabric. I believe the controversy, at its core, concerns the distribution of power and decision-making. Who will be the final arbiter of controversies? Who will control the process, including the selection of convention members, and the limitation, if any, of the State-Convention agenda?[24] In the final analysis, I believe it imperative that proactive steps be taken now to prepare for what will surely confront us later, and, all too possibly, catch

---

and the matter of details was left to the states." *Id.* at 1274.

20. As one author has noted, the State-Convention model under Article V has become little more than a "protest clause." Dixon, *supra* note 16, at 944. In discussing Article V, Professor Dixon further posits that it is "understandable that the convention device has never been used; piecemeal constitutional revision, which is all the people have ever desired, is more expeditiously handled by congressional initiation." *Id.*

21. *See Amendment of the Constitution by the Convention Method Under Article V*, *supra* note 19, at 2.

22. *See* RUSSELL L. CAPLAN, CONSTITUTIONAL BRINKMANSHIP: AMENDING THE CONSTITUTION BY NATIONAL CONVENTION 76, 83 (1988); Ronnie Ellis, *State Lawmakers Call for a Constitutional Convention*, THE DAILY INDEPENDENT (Feb. 22, 2011), http://dailyindependent.com/local/x62852616/State-lawmakers-call-for-constitutional-convention; Christian Gomez, *Texas State Senate Calls for Con-Con*, NEW AMERICAN (Feb. 28, 2011, 8:52 AM), http://www.thenewamerican.com/index.php/usnews/politics/6474-texas-state-senate-calls-for-con-con.

23. *See* Stern, *supra* note 7, at 766–67.

24. *See* Manne, *supra* note 3, at 145. Manne writes that Madison himself worried about the procedural matters relating to Article V. *Id.* ("Madison did not object to the provision for a convention, but noted the problems that might arise over form, quorum, and procedural matters . . . ."). *Id.*

us unaware. If we wait until the moment is upon us, I fear that the necessity of the moment will strike a very bad bargain.

III. PAST PROPOSALS FOR PROACTIVE STEPS

There are few articles of the Constitution as important to the continued viability of our government and nation as Article V.[25]

These twenty-one words introduced one of the most thoughtful and prestigious studies to have evaluated the Article V State-Convention model.[26] As the states had come close on many occasions to forcing Congress to call for a State Convention, the American Bar Association ("ABA") in August 1973 proposed a series of procedures that would delimit the various powers of those individuals and entities most likely to be involved in the State-Convention process.[27] These ABA suggestions, which responded to U.S. Senator Sam Ervin's twice-unsuccessful Senate Bill addressing Article V, continue to provide a very tempered approach to dealing, in advance of necessity, with the State-Convention model.[28]

The ABA Committee was comprised of three judges,[29] a law school dean,[30] a law professor,[31] and various other well-respected lawyers.[32] The Committee formed and considered a variety of questions[33] that, I believe, remain open questions under any Article V analysis, including:

---

25. *Amendment of the Constitution by the Convention Method Under Article V*, *supra* note 19, at 1.

26. *See id.*

27. *See id.* at 5–6.

28. *See id.* at 4. Senator Ervin's proposed bill is contained in Appendix A to the ABA study. *Id.* at 47–57. The ABA Committee went to great lengths to consider, and attempt to improve through their Study, the Ervin proposal. *See id.*

29. *See id.* at iii–iv. The Honorable Sarah T. Hughes, a United States District Judge in Dallas, Texas, oft remembered as the judge that swore in Lyndon B. Johnson on an airplane in Dallas, Texas, after the death of President John F. Kennedy, was the lone federal judicial representative. The Honorable William S. Thompson, a Superior Court Judge from the District of Columbia, and the Honorable C. Clyde Atkins, United States District Judge were the other judicial representatives. *See id.* at iii–iv, ix–x.

30. Dean Albert M. Sacks was the Dean of Harvard Law School at the time. *See id.* at iii.

31. Professor David Dow of the University of Nebraska College of Law was previously the Dean of Nebraska's College of Law. *See id.* at iii, ix–x.

32. The remaining panel members included: Warren Christopher, Esq., of Los Angeles, California; John D. Feerick, Esq., of New York, New York; Adrian M. Foley, Jr., Esq., of Newark, New Jersey; and, Samuel W. Witwer, Esq., of Chicago, Illinois. *See id.* at iii–iv.

33. *See id.* at ix.

> 1) If the legislatures of two-thirds of the states apply for a convention limited to a specific matter, must Congress call such a convention?;
> 2) If a convention is called, is the limitation binding on the convention?;
> 3) What constitutes a valid application which Congress must count and who is to judge its validity?;
> 4) What is the length of time in which applications for a convention will be counted?;
> 5) How much power does Congress have as to the scope of a convention? As to procedures such as the selection of delegates? As to the voting requirements at a convention? As to refusing to submit to the states for ratification the product of a convention?;
> 6) What are the roles of the President and state governors in the amending process?;
> 7) Can a state legislature withdraw an application for a convention once it has been submitted to Congress or rescind a previous ratification of a proposed amendment or a previous rejection?;
> 8) Are issues arising in the convention process justiciable?;
> 9) Who is to decide questions of ratification?[34]

These are just a few of the more pressing questions that will eventually require resolution when the Article V State-Convention method is finally utilized, if ever it is.[35] One thing is certain: there will be tension between the individual states and national government. If there had been consensus on an issue, it would have been addressed by the national government and, more precisely, by Congress without the need for intervention by two-thirds of the States. The purpose of including the State-Convention method in Article V is to provide the individual states with the power to amend when the national government refuses to act.[36] Thus, if we ever reach this point,

---

34. *Id.* at 5.

35. *See* McCleskey, *supra* note 12, at 1003. In addition to the questions posed by the ABA Committee Study, Professor McCleskey notes the "many detailed questions of concern" that will inevitably arise under the State-Convention method, including:

> May a governor exercise his veto power to block legislative ratification? Who decides how a [State] convention is to be created and organized? May a popular vote be substituted for ratification by legislative or convention action? . . . Are there limits on the subjects that may properly be dealt with by amendment? May Congress incorporate a time limit for consideration in a proposed amendment? Does the President have any formal role in the process of initiating proposals?

Id.

36. *See* Gunther, *supra* note 3, at 17. Professor Gunther explained that what he thought "[t]he framers had primarily in mind, then, was that the states should have an opportunity to initiate the constitutional revision process if Congress became unresponsive, arrogant and tyrannical." *Id.* Professor Gunther later underscores this point by stating that "[i]f the state-initiated method for amending the Constitution was designed for anything, it was designed to *minimize* the role of Congress." *Id.* at 23.

we should expect a fierce power struggle between the state and national governments.

Textually, Article V has all the necessary ingredients for a perfect constitutional storm.[37] The textual language implies a power-sharing arrangement, but the most important details—including the issue of judicial review—are notably absent.[38] Furthermore, depending on the issue serving as impetus for the State-Convention method, one cannot anticipate whether the federal courts will intervene or whether the executive will try to participate.[39] Senator Ervin, and subsequently Senator Helms, both attempted during the 1970s to exercise "jurisdiction stripping" of Article V assessments from any court, state or federal, in their proposed Senate bills.[40] It is unclear whether, even without a "jurisdiction-stripping" provision, any congressional attempt to proactively delineate the parameters of Article V would qualify as a non-justiciable political question.

Accordingly, we should follow the lead of the ABA Committee and Senators Ervin and Helms and recommend that some standards be enacted to safeguard the essence of Article V's mandate.[41] The text, standing alone, is ambiguous and vulnerable to manipulation. Unlike Stern, I do not believe there are sufficient safeguards in place.

IV. TEXTUAL LIMITATIONS – AMBIGUITY IN SEARCH OF RESOLUTION

The logical starting point in this endeavor is the text of Article V, for its deceptively plain language conceals an array of ambiguities.[42]

Professor McCleskey accurately depicts Article V as "deceptively plain" and yet, simultaneously ambiguous.[43] The words seem clear enough, but the text is rife with uncertainty. Numerous unanswered questions remain. Madison himself recorded concern in his personal notes: How was a State Convention to be formed? By what rule, or rules, would decisions be reached? What would be the force of State Convention's acts?[44]

---

37. *See* Manne, *supra* note 3, at 135. Manne predicts that "[a] general article V convention, more than any other event possible in our political system, has the potential for a complete reworking of the rules by which government exercises its power." *Id.*

38. *See id.* ("For the constitutional law scholar, the consideration of the convention alternative is a foray into conjecture and speculation. The method has never been used, and the text, history and policy considerations relating to the convention method are all less than unambiguous").

39. *Cf. Bush v. Gore*, 531 U.S. 98 (2000).

40. *See* Gunther, *supra* note 3, at 21–22. Professor Gunther described the "jurisdiction stripping" component of the "Ervin-Helms" proposals to be filled with "grave constitutional doubts." *Id.* at 21.

41. *See id.*

42. McCleskey, *supra* note 12, at 1003.

43. *Id.*

44. *See* Manne, *supra* note 3, at 144–45.

Additionally, "[t]here was no mention of the scope of applications or conventions at any time during debate over the amendment process at the Philadelphia Convention. Contempor[aneous] commentary on these issues was noticeably lacking."[45]

One thing is clear, however, from Article V's text: the design suggests an equal opportunity between the state and national governments for the *proposing* of constitutional amendments.[46] Beyond that, however, there is little textual guidance provided.[47] Congress, ultimately, is a necessary participant in any State-Convention process on at least two levels: first, in determining that the criteria have been satisfied requiring the calling of a Constitutional Convention and, second, in determining whether any proposals deemed to conform with Article V should be submitted to ratification by state legislative vote or further convention.[48]

These are the two textual roles that the national Congress must fill. However, as the ABA Committee appreciates, there are innumerable opportunities for mischief in filling those textually commanded roles.[49] The best occasion for curtailing any politically or constitutionally damaging machinations is in advance of any crisis.[50] If we wait until the crisis is upon us, the various players, both local and national, will be motivated more by the power struggle at hand and less by the constitutional mandates envisioned under Article V.

Stern suggests that while "the text is silent as to what amendments the convention may propose" among other details, we can simply turn to logic to infer solutions.[51] He deftly explains that the textual language provides clues, or expectations, as to what should occur upon the calling of an Article V Convention.[52] My concern is that Stern proffers only one interpretation of language that is anything but historically clear.[53] For each

---

45. *Id.* at 146.

46. *See* Martig, *supra* note 9, at 1258–61.

47. *See* Gunther, *supra* note 3, at 23 ("Congress was only given two responsibilities under [the State Convention] portion of Article V" and, he believed, "that, properly construed, these are extremely narrow responsibilities.").

48. *See id.*

49. *See id.*

50. *See id.* at 24 (calling on Congress to address these issues that are "long overdue"). Professor Gunther further opines that: "If there is merit to my tale of confusion and uncertainty, Congress surely owes it to the country to consider the differing views about Article V and to clarify the misimpressions under which so many state legislatures may have [already] acted." *Id.*

51. Stern, *supra* note 7, at 772.

52. *See id.* at 774–75 (discussing Michael Stokes Paulsen, *A General Theory of Article V: The Constitutional Lessons of the Twenty-Seventh Amendment*, 103 YALE L.J. 677, 738, 740 (1993)).

53. *See* Manne, *supra* note 3, at 148 (observing that "[t]he most striking thing about this scholarly analysis of history, frequently couched in terms of the 'intent of the framers' . . . is not that there should be such irreconcilable disagreement, but that there should be this

article that suggests that an Article V State Convention poses no threat to our constitutional integrity and that the "solutions" to such Convention are discernable, there is another warning of the perils inherent in the State-Convention model.[54]

While I will refrain from entering the debate as to whether the calling of an Article V State Convention is prudential or risky, I do believe that pressing forward without some textual resolution or instruction, such as a legislative mandate or judicial pronouncement, will yield a troubled State-Convention process. The ABA Committee's Study referenced above provides an exceptional opportunity to revisit this longstanding issue and work toward implementable solutions.[55] The debates that continue have not provided tangible solutions and I am not convinced that these writings clearly define the outstanding issues. Instead, our writings are academic, in the purest and most literal sense. They are predictive. They are historical. They are even entertaining.[56] While these writings might be influential to those ultimately called upon to resolve the Article V issue, I would welcome a change of focus from debating what we are certain will occur, based on textual expectations or interpretations, to a call for action demanding counsel as to how to conduct ourselves when the State-Convention moment is finally upon us.[57]

The text itself is neither dispositive nor necessarily informative in the constitutional sense. If the devil is in the details, then it is little wonder that the text of Article V bedevils all who strive to distill its true meaning and predict its future application.

---

proliferation of [opposing] statements of what history 'clearly demonstrates' at all").

54. *See id.* at 146–47 (cataloguing the varying positions of scholars).

55. *But see* Charles L. Black, Jr., *Amending the Constitution: A Letter to a Congressman*, 82 YALE L.J. 189, 189 (1972) (calling the development of proactive legislation similar to the failed bills advanced by Senators Ervin and Helms in the 1970s as "a national calamity").

56. *See* Manne, *supra* note 3, at 149–56. Perhaps the most entertaining presentation of the Article V quagmire is Manne's "Play in One Act" entitled "The Second Coming of Thomas Jefferson." In this play, set forth within his larger article, Manne posits a ghostly meeting between Professor Charles L. Black and Thomas Jefferson to discern the deeper meanings of Article V.

57. *But see* Black, *supra* note 55, at 194. Admittedly, Professor Black takes the exact opposite position. Professor Black would counsel against any proactive legislation noting that "[i]t is most unwise to try to settle such questions at a time when national attention is not and cannot be keenly focused upon them, and intense national debate be thus generated." *Id.* Unlike the approach recommended in the current article, Black would rather wait for the crisis, constitutional though it may become, and allow the attendant focused debate and national attention to moderate behavior. My thesis differs in that I fear such a situation would likely yield emotional solutions as opposed to rational decisions devoid of emotion.

V. CURRENT "SAFEGUARDS" – SUPREME CONFUSION?

> *The question of justiciability pervades the discussion of all the above issues. How these controversies are ultimately resolved may depend more on who decides them than on the textual considerations just discussed or the historical and policy considerations that follow. Article V is silent on the issue of justiciability. Absent a clear grant or denial of judicial review, the question of justiciability is, by its nature, a policy issue. One should note that, in theory at least, each Article V question presents a separate issue as to justiciability; some might be held justiciable while others might not.*[58]

Issues surrounding Article V and the State-Convention method are not completely foreign to the United States Supreme Court. In fact, former Chief Justice Rehnquist provided tangential insight into the Article V process as recently as 1989.[59] As we move toward an eventual confrontation with Article V's State-Convention model, we must be forward thinking in our desire to add clarity to the process. Rather than remain comfortable debating the potential breadth of any State Convention, scholars and legislators should join forces to provide a solution to what remains an obvious quandary. The guidance provided by the Supreme Court is neither definitive nor clear and only adds to the lingering confusion. Will this be an area where, ultimately, the Court will "say what the law is?"[60]

In the 1920s, the Supreme Court had two occasions to decipher Article V's amendment process. In 1920, the Court presented a thorough interpretation of Article V's ratification provisions in *Hawke v. Smith*.[61] The discrete question presented was whether the state of Ohio could, through State Constitutional provision, mandate a voter referendum for ratification purposes.[62] In finding the voter ratification method in conflict with Article V, Justice Day explained that "[t]he act of ratification by the state derives its authority from the federal Constitution to which the state and its people have alike assented."[63] The Court also reminded, "[i]t is not the function of courts or legislative bodies, national or state, to alter the method which the Constitution has fixed."[64]

---

58. Manne, *supra* note 3, at 141–42.

59. *See* Uhler v. Am. Fed. of Labor-Cong. of Indus. Org., 468 U.S. 1310, 1312 (1989).

60. *See* Marbury v. Madison, 5 U.S. 137, 177 (1803). While *Marbury* certainly established the notion of judicial review, other principles, including the concept of justiciability and the political question doctrine, often operate to curtail the power of judicial review.

61. *See generally* Hawke v. Smith, 253 U.S. 221 (1920).

62. *See id.* at 231.

63. *Id.* at 230.

64. *Id.* at 227.

Accordingly, the Court struck down the Ohio attempt at voter ratification finding that the Framers did not intend for the people, acting individually as such, to play any role in Article V's ratification process.[65] Rather,

> The fifth article is a grant of authority by the people to Congress. The determination of the method of ratification is the exercise of a national power specifically granted by the Constitution; that power is conferred upon Congress, and is limited to two methods, by action of the Legislatures of three-fourths of the states, or conventions in a like number of states. The framers . . . might have adopted a different method. Ratification might have been left to a vote of the people, or to some authority of government other than that selected. The language of the article [on this point] is plain, and admits of no doubt in its interpretation.[66]

Thus, the first case directly addressing the role of the people in Article V found that the historical meaning of "legislatures" was clear.[67] Legislatures, when used in the Constitution, mean those elected officials that serve as representatives of the people.[68] As Justice Day reminds, when the framers "intended that direct action by the people should be had they were no less accurate in the use of apt phraseology to carry out such purpose."[69]

What does *Hawke* add to this discussion? The Court had no trouble reaching a merits-based decision regarding Article V. In fact, the Court was willing to tell "the people" that their legislatures could, in fact, bind them to decisions without requiring any direct input from their citizens. Ohioans, as a people, were displeased that their legislature was capable of binding them to the Eighteenth Amendment. But, *Hawke* had little difficulty sanctioning the legislative ratification process.[70] Article V's ratification provisions were deemed to be clear, unambiguous and, most importantly, to exclude any role of the people, individually or directly.[71] We can interpret *Hawke* as allowing the Court to resolve Article V power struggles between state citizens and their state legislative representatives. Further, *Hawke* underscores that at least some provisions within Article V are subject to judicial review.

Two years later, the Court addressed the question of whether the Nineteenth Amendment had become part of the federal Constitution in

---

65. *See id.*
66. *Id.* (internal citation omitted).
67. *See id.*
68. *See id.* at 227.
69. *Id.* at 228.
70. *See id.*
71. *See id.*

Leser v. Garnett.[72] Maryland sought to disqualify Cecilia Streett Waters and Mary D. Randolph from exercising their right of suffrage recently guaranteed by the Nineteenth Amendment.[73] In presenting its case to the Court, Maryland first suggested that the substantive requirements of the Nineteenth Amendment, giving women the right to vote, would, "if made without the State's consent, destroy[] its autonomy as a political body."[74] Justice Brandeis and the Court reminded Maryland that its previous refusal to ratify the Fifteenth Amendment, providing black males the right to vote, also did not prevent that Amendment from taking effect in every state, including Maryland.[75] Maryland's second challenge, like that presented in *Hawke*, was that because of numerous state constitutional prohibitions, state legislatures are prohibited from acting for the people in ratifying amendments to the Constitution.[76] Justice Brandeis, much like Justice Day, reminded that:

> [T]he function of a state Legislature in ratifying a proposed amendment to the federal Constitution, like the function of Congress in proposing the amendment, is a federal function derived from the federal Constitution; and it transcends any limitation sought to be imposed by the people of a state.[77]

Hence, *Leser*, like *Hawke*, reminds that the Supreme Court may invoke its power to enforce Article V's provisions on the people of an objecting state. A merits-based review of Article V is certainly possible for any State-Convention question. *Leser* and *Hawke* also appear to be cases upholding Article V's structural components. If a reviewing court finds that questions relating to the State-Convention model are structural, or that the Framers' intent on the particular question is clearly discernable, there is little doubt that the Court has the power, precedentially as well as jurisprudentially, to resolve the controversy. However, neither *Leser* nor *Hawke* lends any insight into a situation where the question is not as clearly presented or as clearly defined.

The most troubling case casting its shadow upon Article V is *Coleman v. Miller.*[78] *Coleman* considered the Supreme Court's ability to intervene in the amendment process under Article V where Congress had acted.[79] In reaching its decision that the issue of whether a State's ratification of a

---

72. *See* 258 U.S. 130, 136 (1922).

73. *See id.* at 135–36.

74. *Id.* at 136.

75. *See id.* "That the Fifteenth is valid, although rejected by six states including Maryland, has been recognized and acted on for half a century." *Id.* (citations omitted).

76. *See id.* at 136–37.

77. *Id.* at 137.

78. 307 U.S. 433 (1939).

79. *See id.*

Congressionally proposed amendment[80] was valid, despite an individual state's initial rejection and subsequent ratification, the Court announced that:

> [W]e think that the Congress in controlling the promulgation of the adoption of a constitutional amendment has the final determination of the question whether by lapse of time its proposal of the amendment had lost its vitality prior to the required ratifications.[81]

While the Court refused to rule directly on whether the various Article V issues constituted non-justiciable political questions, having been divided on the issue,[82] the decision is replete with language deferring to the national Congress.[83] Four Justices agreed that the issues presented were non-justiciable political questions, but failed to command a majority view.[84] Thus, there is uncertainty as to whether a modern court would be receptive to resolving an Article V challenge where the claim involves the role of Congress, like that addressed in *Coleman*,[85] or would find that such question was a non-justiciable political question. While *Leser* and *Hawke* contemplate an active decisional role for the Court,[86] *Coleman* creates an element of doubt in the equation.[87]

Over fifty years later, Chief Justice Rehnquist, sitting as Circuit Justice, refused to issue a stay intervening in the California controversy as to

---

80. *Coleman* involved the State of Kansas' initial rejection and subsequent ratification of the Child Labor Amendment. *Id.* at 435–36.

81. *Id.* at 456.

82. *See id.* at 447 ("Whether this contention presents a justiciable controversy, or a question which is political in its nature and hence not justiciable, is a question upon which the Court is equally divided and therefore the Court expresses no opinion on that point.").

83. *Id.* at 450. The Court, deferring completing to the national Congress for resolution of the ratification decision opined:

> We think that in accordance with this historic precedent the question of the efficacy of ratifications by state legislatures, in the light of previous rejection or attempted withdrawal, should be regarded as a political question pertaining to the political departments, with the ultimate authority in the Congress in the exercise of its control over the promulgation of the adoption of the amendment.

Id.

84. *See* Uhler v. Am. Fed. of Labor-Congress of Indust. Orgs., 468 U.S. 1310, 1312 (1989) ("[F]our Justices of the Court adopted the position that the Court lacked jurisdiction to rule on questions arising in connection with the ratification of a constitutional amendment because all such questions were 'political' in nature. But that position did not command a majority in *Coleman* . . . .").

85. *See Coleman*, 397 U.S. at 433.

86. *See* Lesser v. Garnett, 258 U.S. 130 (1922); Hawke v. Smith, 253 U.S. 221 (1920).

87. *Coleman*, 307 U.S. at 433.

whether the voters could require their state legislators to join in the call for an Article V State Convention.[88] The basis for the refusal, set forth in a brief three-page opinion, was that California had provided an independent and adequate state ground for reaching its decision and, therefore, the Supreme Court should not intercede.[89] Justice Rehnquist did, however, provide some illuminating dicta as to the political question doctrine, predicting that neither state nor federal courts would be completely powerless in addressing all Article V issues.[90]

Four Supreme Court opinions, spanning nearly 100 years, have yielded little in the way of resolving the potential Article V controversies that the State-Convention model presents.[91] In fact, most of the questions presented in the ABA Committee Study were not addressed by the Court, as the two most recent decisions were resolved without clearly delineating the role of the states, Congress, or, equally important, the judiciary, in any forthcoming Article V State-Convention process.[92] The soft doctrines of justiciability and political question provide the Court with an opportunity to remove itself from any particular and divisive issue.[93] Precisely because these doctrines are "soft" and malleable, one cannot predict with much certainty that the Court will rebuff, or receive, any particular challenger if the Court deems intervention necessary or the Constitutional issue paramount.[94] Thus, relying on the Court as an arbiter in this arena is speculative and risky.[95] With no clarity in reach, Article V presents the opportunity for Supreme confusion.[96]

---

88. *See Uhler*, 468 U.S. at 1312.

89. *See id.* at 1311.

90. *See id.* at 1312.

91. *See* Manne, *supra* note 3, at 157. "The resolution of the scope issue, and other article V questions as well, may hinge on who decides them." *Id.*

92. The role of the President under Article V was narrowed by the Supreme Court in *Hollingsworth v. Virginia*, 3 U.S. 378 (1798). The *Hollingsworth* holding was expanded from its facts by the Court in *Hawke* which exclaimed that "[a]t an early day this court settled that the submission of a constitutional amendment did not require the action of the President." Hawke v. Smith, 253 U.S. 221, 229 (1920).

93. Of course, the mere existence of these doctrines does not mean the Court will invoke them. *See* Bush v. Gore, 531 U.S. 98 (2000) (where the Court addressed the Florida vote controversy without ever addressing the issues of standing, ripeness, political question, or adequate and independent state grounds).

94. *See id.*

95. The state courts, however, have played a more vibrant role in interpreting Article V's parameters. In 1996, the Oklahoma Supreme Court found that the call for a convention "must come from the Legislature acting freely and without restriction or limitation, [and not directly] from the people through their initiative power." *See In re* Initiative Petition No. 364, 930 P.2d 186 (Okla. 1996). In doing so, the Oklahoma Supreme Court struck down the voter-driven initiative to instruct the Oklahoma legislature to call for a federal constitutional convention. *Id.* at 191. The case involved a proposed initiative that would have instructed the legislature to call for a national constitutional convention addressing legislative term limits.

V. CONCLUSION

Everything I have said constitutes conjecture about the past and advice about the future.[97]

Professor Gerald Gunther uttered these words delivering a speech in 1979.[98] The sentiments, though not my own, are as potent a closing vehicle for me as they were for Professor Gunther over thirty years ago. Humility requires that I confess my writing, though well intended and, hopefully, equally well researched, is as speculative as any other on the topic of Article V. Rather than press forward in the certainty that any entity will adopt my interpretation on Article V, I can only hope to give advice as to how future actors should address the topic.

Much like Professor Gunther and the ABA Committee, I urge Congress to be proactive. Legislation defining the parameters of an Article V State-Convention scenario should be considered now, well in advance of any need or crisis mandating a particular response to a particular topic.[99] The worst-case scenario, in my opinion, is not finding ourselves without an answer as to whether any State Convention under Article V will be limited or unlimited, but rather finding ourselves face to face with an Article V State Convention without procedures specifying the roles of the various players. The value in proactively establishing convention procedures is that such advance directives should eliminate tying any power struggle with a particular topic, such as abortion, school prayer, or a federal balanced budget amendment. Procedures enacted in a hostile environment will be influenced as much by emotion as by reason.

The template needed for such proactive legislation is already in draft form, though substantial reconsideration must be given to the failed bills of Senators Ervin and Helms.[100] Therefore, I would recommend that Congress create an Article V sub-committee, with equal House and Senate

---

The Supreme Court of Arkansas and the United States District Court of Maine followed this same approach in 1996 and 1997, respectively. *See* Donovan v. Priest, 931 S.W.2d 119 (Ark. 1996); League of Women Voters of Me. v. Gwadosky, 966 F. Supp. 52 (Me. 1997).

96. *See* Manne, *supra* note 3, at 157 ("[A]ny argument regarding justiciability, no matter how well reasoned, can be made instantly wrong by five Justices.").

97. Gunther, *supra* note 3, at 25.

98. *Id.* at 1. Professor Gunther, who at the time was teaching at Stanford University's Law School, delivered the John A. Sibley Lecture in Law at the University of Georgia School of Law on May 24, 1979, and his comments were ultimately set forth in an essay format by the Law Review. *Id.*

99. *But see* Black, *supra* note 55, at 195 ("These problems can and should be solved when they arrive, by the Congress empowered to solve them, and on the basis of all the factors now unknowable and then existing."). Black's approach assumes a fact that this author neither concedes nor embraces: that the parameters of Article V will be for Congress, rather than the courts, state legislatures or executive, to decide.

100. *See* Gunther, *supra* note 3, at 21.

membership, to work in concert with the American Law Institute and the American Bar Association to create balanced and moderate procedures that would govern any Article V State Convention. Much like the previous ABA Committee,[101] a modern committee should include jurists, academics, and lawyers of the highest quality and experience. Additionally, the American Law Institute has proven itself adept at facing and proposing considered resolution for tough legal issues.[102]

Space prevents me from setting forth a detailed proposal of what items should be considered or how such sub-committee would function. But, then again, perhaps the details are best left to a sub-committee once created. While I would not confine any potential Article V sub-committee to a discrete series of questions for resolution, I would be remiss if I did not suggest that the 1973 ABA Committee's Study[103] provides an excellent template of where to begin. The consideration of those questions evaluated by the ABA Committee in 1973,[104] all still being unresolved, provides an excellent starting place.

In the end, I confess to caring less about the detail of what items are considered in any particular order. Instead, I am more concerned that a deliberate, proactive consideration of the relevant procedural issues takes place well in advance of any Article V State Convention. The importance of having sound procedures in place prior to the invocation of Article V assures all that the power struggle between state and national governments the framers sought to avoid actually can be avoided. Despite Stern's confidence in our government to adequately confront an Article V State Convention, I fear any bargain borne out of necessity.

The time for action is now. For if there is one truism of constitutional law it is that the textual vacuum of a constitutional provision will be filled by something.[105] I harbor grave concerns about not only *what* will fill the Article V void, but also *who* will ultimately make that decision.

Benjamin Franklin was a very wise man. Necessity, as we know, never made a good bargain.[106] Let us avoid that necessity and strive to fill the Article V void now. If we wait, having literally had centuries to resolve these issues, we deserve the "bargain" that befalls us.

---

101. *See Amendment of the Constitution by the Convention Method Under Article V*, *supra* note 19, at 1.

102. *See id.*

103. *See id.*

104. *See id.*

105. *See, e.g.*, McCulloch v. Maryland, 17 U.S. 316 (1819).

106. FRANKLIN, *supra* note 1, at 10.

A BRIEF REPLY TO PROFESSOR PENROSE

MICHAEL STERN*

"It's tough to make predictions, especially about the future."[1] In her thoughtful and gracious response, Professor Mary Margaret Penrose emphasizes the uncertainty inherent in the Article V Convention process.[2] Indeed, there are a number of unsettled issues surrounding that process, ranging from administrative matters, such as where the convention will meet, to more complex and controversial questions, such as how it will vote.[3] I agree with her that there are elements of uncertainty in the convention process and that resolving issues in advance of a convention is a laudable goal.

Uncertainty about some things, however, does not mean uncertainty about everything. Take, for example, Senator Ferry's 1869 warning, quoted by Penrose, that a convention cannot be limited to "'the simple amendment which you are proposing to it. It may go on to amend your State constitution and to subvert the whole machinery of your State government, and there is no power in your State to stop it.'"[4] This is certainly wrong. In the first place, an Article V Convention cannot by itself amend anything; its authority is limited to proposing amendments.[5] Second, it can only propose amendments to the federal Constitution, not to state constitutions.[6]

Moreover, while it is true that there has never been an Article V Convention to propose amendments,[7] it does not follow that there is no experience from which reasonable predictions can be made about the Article V process. In fact, while an Article V Convention cannot propose amendments to state constitutions, there have been many state

---

* The author formerly served as Senior Counsel in the Office of General Counsel, U.S. House of Representatives, as Deputy Staff Director for Investigations for the Senate Committee on Homeland Security and Governmental Affairs, and as Special Counsel to the House Permanent Select Committee on Intelligence. He writes about constitutional issues relating to the legislative process and other congressional legal issues at www.pointoforder.com.

1. ANTHONY ST. PETER, THE GREATEST QUOTATIONS OF ALL-TIME 264 (2010).

2. *See generally* Mary Margaret Penrose, *Conventional Wisdom: Acknowledging Uncertainty in the Unknown*, 78 TENN. L. REV. 789 (2011) (responding to Michael Stern, *Reopening the Constitutional Road to Reform: Toward a Safeguarded Article V Convention*, 78 TENN. L. REV. 765 (2011)).

3. *See* Penrose, *supra* note 2, at 794–95.

4. *Id.* at 794 n.19 (quoting Ralph R. Martig, *Amending the Constitution, Article Five: The Keystone of the Arch*, 35 MICH. L. REV. 1252, 1272 (1936)).

5. *See* U.S. CONST. art. V.

6. *See id.*

7. *See* Penrose, *supra* note 2, at 790.

constitutional conventions throughout American history that have exercised precisely that power.[8] A recent work identifies some 233 such state conventions that "have much in common in the way their delegates have been selected, their business conducted, and their work submitted."[9] The history of these conventions reveals no tendency for them to fall under the control of radical or irresponsible elements and should provide some comfort that there is nothing inherently dangerous about such proceedings.[10]

This is in no way to disagree with Penrose's suggestion that the planning for an Article V Convention begin now.[11] It is important, however, to differentiate between two distinct objectives of such an effort. The first objective is to address collective action problems, which present a barrier to holding a convention. Absent an agreement on the location, funding, organization, and procedures of a convention, such a convention may disintegrate before it ever gets started, making it impossible to address the substantive constitutional amendment that the states wish it to consider.

The second objective would be to prevent a runaway convention, which was the focus of my original article.[12] It must be emphasized that the collective action problems noted above do not make a runaway convention more likely; to the contrary, they make it less likely that the convention will be able to consider or propose any amendment at all. The risk of a failed or stillborn convention seems to me to be significantly larger than that of a runaway convention.

Nevertheless, my article proposes a number of additional safeguards that could be adopted to minimize any risk of a runaway convention.[13] The safeguards, though perhaps unnecessary, would serve as confidence-builders to assure state legislatures that an Article V Convention will not be hijacked in the service of an unknown agenda.[14] Enacting some or all of these measures would be consistent with Penrose's approach, and she does not appear to dispute that they would be efficacious.[15]

Thus, although Professor Penrose and I view this issue from divergent perspectives, the practical differences in our approaches may be fairly modest. There are two areas, however, where our disagreements deserve further explication.

1. The Role of the States versus Congress. Penrose aptly notes that "[t]he purpose of including the State Convention method in Article V is to

---

8. *See* JOHN J. DINAN, THE AMERICAN STATE CONSTITUTIONAL TRADITION 7 (2009).

9. *Id.* at 12.

10. *See generally id.*

11. *See* Penrose, *supra* note 2, at 793–94.

12. *See generally* Michael Stern, *Reopening the Constitutional Road to Reform: Toward a Safeguarded Article V Convention*, 78 TENN. L. REV. 765 (2011).

13. *See id.* at 784–87.

14. *See id.* at 787–88.

15. *See generally* Penrose, *supra* note 2.

provide the individual states with the power to amend when the national government refuses to act."[16] Consistent with that purpose, Article V gives Congress an extremely limited role in the convention process.[17] Congress has the ministerial duty of calling the convention when the requisite applications have been received, and it must determine the method of ratification for any amendment proposed by the convention.[18]

Penrose nonetheless proposes that Congress consider and enact legislation to establish procedures to govern a convention.[19] While it is arguable that Congress has some limited authority in this regard, such as determining the place of the convention, any attempt by Congress to prescribe the rules of the convention would be in considerable tension with the purposes of Article V.

Commentators have concluded that Congress therefore may not exercise any discretionary authority over the convention method of amendment.[20] Professor William Van Alstyne, for example, finds that "Congress [is] supposed to be mere clerk of the process convoking state-called conventions."[21] Professor Robert Natelson contends that Congress is supposed to act as the agent of the state legislatures with respect to the calling of an Article V Convention.[22] As an agent, Congress must follow the directions received from its principals and "may not impose rules of its own on the states or on the convention."[23]

The task of setting rules and procedures for an Article V Convention therefore must fall to the states rather than to Congress. Congress can play a limited, but important, role by declaring in advance how it will count applications for a limited convention and whether it will recognize proposed amendments that exceed the scope of a limited convention. Any attempt by Congress, however, to impose rules and procedures on a convention would be constitutionally questionable at best and would give rise to the very type of state-federal imbroglio that Penrose fears.[24]

---

16. *Id.* at 795.

17. *See* U.S. CONST. art. V.

18. *See id.*

19. *See* Penrose, *supra* note 2, at 797–98.

20. *See, e.g.*, William W. Van Alstyne, *Does Article V Restrict the States to Calling Unlimited Conventions Only?—A Letter to a Colleague*, 1978 DUKE L.J. 1295 (1978).

21. *Id.* at 1303.

22. *See* Robert. G. Natelson, *Amending the Constitution by Convention: A Complete View of the Founders' Plan (Part 1 in a 3 Part Series)*, POLICY REPORT NO. 241 (GOLDWATER INSTITUTE), Sept. 2010, at 19–22, *available at* http://www.goldwaterinstitute.org/article/5005.

23. *Id.* at 21.

24. *See* Penrose, *supra* note 2, at 793–94. On the other hand, it is certainly permissible for Congress to hold hearings on an Article V Convention, to hear from experts about various issues that may arise with regard to a convention, and even to issue recommendations on rules and procedures that either the states, the convention, or both may wish to adopt.

A state-led effort to prescribe rules and procedures for an Article V Convention would no doubt be more difficult and cumbersome, but it is far from impossible. A uniform act to establish the required rules could be developed, and each state that enacted the law would instruct its delegates to vote for such rules as the first act of the convention. Comparable efforts to address state functions under the federal Constitution with respect to the electoral college have enjoyed considerable success.[25] Moreover, the state conventions that ratified the Twenty-first Amendment relied heavily on a uniform state law approach, with many states adopting a prototype statute verbatim.[26] A uniform state law approach to the collective action problems of an Article V Convention is thus entirely feasible.

2. The Role of the Courts. Much of Penrose's analysis is devoted to showing that the courts cannot be relied upon to resolve any particular controversy regarding the Article V Convention method of amendment.[27] This is true both because there is little case law regarding the merits of most of the potential areas of legal disagreement, and because doctrines of justiciability make it uncertain when, if ever, the courts might reach the merits of any such disagreement.[28]

Penrose is certainly correct in this regard. However, I have difficulty accepting the implication that this state of affairs somehow makes the Article V Convention a riskier proposition. The proposition appears to rest on the assumption that a legislative or constitutional process is inherently unpredictable unless the courts have pronounced how it is to operate. In fact, our normal legislative processes operate based on rules which are set by the legislative body or directly by the state or federal constitution, with little or no judicial intervention.[29]

Nor is it the case that court decisions, generally speaking, necessarily make the law more clear or predictable. The opening observation about predicting the future is at least as applicable to judicial decision-making as to baseball games or other areas of human endeavor.

---

25. For example, the National Conference of Commissioners on Uniform State Laws produced in 2010 a Uniform Faithful Presidential Electors Act. *See generally* NAT'L CONFERENCE OF COMM'RS ON UNIF. STATE LAWS, UNIFORM FAITHFUL PRESIDENTIAL ELECTORS ACT (2010) *available at* http://www.law.upenn.edu/bll/archives/ulc/fpe/2010_final.htm. Additionally, the National Popular Vote bill, which would commit enacting states to award their presidential electors to the candidate receiving a majority of the national popular vote, has been enacted in six states and the District of Columbia. *See* Explanation of the National Popular Vote, NATIONAL POPULAR VOTE, http://www.nationalpopularvote.com/pages/explanation.php (last visited May 5, 2011).

26. *See* Richard M. Evans, *How Alcohol Prohibition Was Ended*, http://www.drug library.org/think/~jnr/endprohb.htm (last visited May 5, 2011).

27. *See* Penrose, *supra* note 2, at 799–803.

28. *See id.*

29. *See generally* United States v. Ballin, 144 U.S. 1 (1892); Marshall Field & Co. v. Clark, 143 U.S. 649 (1892).

So long as there is a doubt as to the constitutional validity of an out-of-scope amendment, it is highly unlikely that an Article V Convention would propose such an amendment. Doing so would entail substantial legal and political risks for delegates who have been instructed by their states to limit consideration to a particular amendment or subject. It would also mean that the convention's work would likely be for naught, as it could be nullified by Congress or the state legislatures themselves.

The possibility of judicial review merely adds an additional "veto point" to the process, and therefore makes it even more unlikely that the convention would propose an out-of-scope amendment. While it is uncertain whether the courts would reach the merits of a challenge to the constitutionality of the amendment and, if so, what they would ultimately decide, it can be said with confidence that there would be legal challenges that would add more time and expense to an already complex ratification process.

Thus, far from enhancing the risk of an Article V Convention, the possibility of judicial review contributes to the stability of the process. If the states set forth guidelines for the conduct of a convention, including a limit on the amendments the convention may consider, the safe harbor will be for the convention to stay within those guidelines. The same is true for delegates who are given instructions by their states. Violating the guidelines or disobeying the instructions would certainly lead to protracted and expensive litigation and could potentially result in invalidation of a proposed amendment, on the one hand, and civil or criminal sanctions for the faithless delegate, on the other. There will be a strong incentive, therefore, for delegates individually and the convention as a whole to turn square corners in implementing the mandate from the states.

Life is full of uncertainty and comes with few guarantees, other than death and taxes. Predictions are hard, especially about the future. Yet the Framers well understood these facts when they chose to include the convention method of amendment within Article V. Specifically, they understood that there were no guarantees that the federal government they established would be restrained by mere "parchment barriers." To address this uncertainty, they designed the Article V Convention as a means for the states to resist federal encroachment. The process they established is as yet unused, but it is not unsafe.[30] With appropriate action by the states, it can be made safer still.

---

30. *See* PAUL J. WEBER & BARBARA A. PERRY, UNFOUNDED FEARS: MYTHS AND REALITIES OF A CONSTITUTIONAL CONVENTION 13–14 (Paul L. Murphy ed. 1989). "[W]hat the Founders did was far more cautious, careful and respectful of citizens' rights and established procedures than the term 'runaway convention' implies." *Id.* at 13.

THE CASE FOR THE REPEAL AMENDMENT

RANDY E. BARNETT*

Today, a political movement has arisen to oppose what seems to be a highly discretionary and legally unconstrained federal government.[1] Beginning in the Bush Administration during the Panic of 2008 and accelerating during the Obama Administration, the federal government has bailed out or taken over banks, car companies,[2] and student loans.[3] It is now preparing to vastly expand the Internal Revenue Service to help it take charge of the practice of medicine for the first time in American history.[4]

This marked and rapid increase of power has shaken many Americans who are now looking to the United States Constitution with renewed interest in the limits it imposes on the powers of Congress. Despite what the Constitution says, however, federal judges have allowed Congress to exceed its enumerated powers for so long, it seems they no longer entertain even the possibility of enforcing the text.[5]

---

* Carmack Waterhouse Professor of Legal Theory, Georgetown University Law Center. I wish to thank Anastasia Killian for her research assistance. Permission for instructors to distribute this essay for educational purposes is hereby granted.

1. *See, e.g.*, TEA PARTY PATRIOTS, *Mission Statement and Core Values*, http://www.teapartypatriots.org/Mission.aspx (last visited May 5, 2011) (noting that the "impetus for the Tea Party movement is excessive government spending and taxation"). The Tea Party is not the only political movement opposing the unchecked growth of federal power, but it is one of the most visible and influential. *See* Randy E. Barnett, *The Tea Party, the Constitution, and the Repeal Amendment*, 105 NW. U. L. REV. Colloquy 281 (2011), http://www.law.northwestern.edu/lawreview/colloquy/2011/10.

2. *See* Emergency Economic Stabilization Act of 2008, Pub. L. No. 110-343, §§ 101–136, 122 Stat. 3765, 3767–3800 (codified as amended in 12 U.S.C. §§ 5211–41) (creating the Troubled Assets Relief Program ("TARP")). Companies receiving TARP money, to name only a few, include Citigroup, Bank of America, JP Morgan Chase, Wells Fargo, Goldman Sachs, Morgan Stanley, PNC Financial Services Group, U.S. Bancorp, Capital One Financial, Regions Financial Corporation, SunTrust, GMAC Financial Services, General Motors, and Chrysler. *See Bailout Recipients*, PROPUBLICA.ORG, http://bailout.propublica.org/list/index (last visited May 5, 2011).

3. *See* Health Care and Education Reconciliation Act of 2010, Pub. L. No. 111-152, §§ 2201–13, 124 Stat. 1071–81 (codified as amended in scattered sections of 20 U.S.C.) (ending federal subsidies of student loans, consolidating existing student loans, and granting the federal government authority to issue and oversee future loans).

4. *See id.* § 1002, 124 Stat. 1029, 1032-33 (codified as amended at I.R.C. § 5000A (West 2010)) (imposing penalties on individuals who fail to maintain minimum essential coverage); H. COMM. ON WAYS AND MEANS, 111TH CONG., THE WRONG PRESCRIPTION: DEMOCRATS' HEALTH OVERHAUL DANGEROUSLY EXPANDS IRS AUTHORITY 4, 7–9 (2010) (estimating the "IRS may need to hire as many as 16,500 additional auditors, agents, and other employees" to implement the new Act).

5. *See, e.g.*, Gonzalez v. Raich, 545 U.S. 1 (2005) (holding that Congress has the

Judges appointed by both Republican and Democratic presidents largely operate within what academics call the "New Deal settlement."[6] By this it is meant that the courts allow Congress to exercise unchecked power over the national economy and everything that may affect it, limited only by the express guarantees of the Bill of Rights.[7] In this arena, with some exceptions, the post-New Deal judiciary disagrees only on whether other unenumerated rights may also receive protection and, if so, which ones.[8] But whatever few additional "fundamental" rights may be recognized, they do not include the protection of any so-called "economic liberty" that might inhibit the national regime of economic regulation.[9]

In this manner, the original scheme of islands of federal powers in a sea of liberty has been transformed into a regime of islands of rights in a vast sea of national power.[10] But judicial passivism is not the only cause of expanding congressional power. Also responsible are two changes to the Constitution's structure that were made in 1913 as "populist" or "progressive" reforms but which fundamentally altered the relationship between the federal government, the states, and the people as it appears in the Constitution's text.[11]

The first change was the Sixteenth Amendment.[12] By giving Congress the power to impose an income tax, the amendment allowed Congress to tax, spend, and redistribute income to a degree previously unimaginable. The Sixteenth Amendment has enabled Congress to evade the limits placed

---

authority under the Commerce Clause to prohibit the local cultivation and use of marijuana in accordance with state law). Justice Thomas's dissent states, "If Congress can regulate this under the Commerce Clause, then it can regulate virtually anything—and the Federal Government is no longer one of limited and enumerated powers." *Id.* at 57–58 (Thomas, J., dissenting).

6. *See, e.g.*, Michael C. Dorf, *Legal Indeterminacy and Institutional Design*, 78 N.Y.U. L. REV. 875, 880 (2003); Richard A. Epstein, *The Classical Liberal Alternative to Progressive and Conservative Constitutionalism*, 77 U. CHI. L. REV. 887, 903 (2010); Laura Kalman, *The Constitution, the Supreme Court, and the New Deal*, 110 AM. HIST. REV. 1052, 1066 (2005); Larry Kramer, *The Supreme Court 2000 Term—Foreword: We the Court*, 115 HARV. L. REV. 4, 122 (2001).

7. *See* Kramer, *supra* note 6, at 125.

8. *Compare* Washington v. Glucksberg, 521 U.S. 702, 706 (1997) (declining to recognize physician-assisted suicide as a "fundamental right" protected by the Due Process Clause), *with* Lawrence v. Texas, 539 U.S. 558 (2003) (holding state sodomy law unconstitutional without employing the two-step analysis used in *Glucksberg*).

9. *See* Randy E. Barnett, *Scrutiny Land*, 106 MICH. L. REV. 1479 (2008) (discussing the evolution and operation of the Supreme Court's "fundamental rights" doctrine).

10. *See* STEPHEN MACEDO, THE NEW RIGHT V. THE CONSTITUTION 32 (rev. ed. 1987).

11. *See* GERARD N. MAGLIOCCA, THE TRAGEDY OF WILLIAM JENNINGS BRYAN: CONSTITUTIONAL LAW AND THE POLITICS OF BACKLASH 134 (2011) (locating the Sixteenth and Seventeenth Amendments as originating in the populist movement and adopted by progressives).

12. U.S. CONST. amend. XVI.

on its power by funding all sorts of activities not otherwise within its enumerated powers—a proposition that the Supreme Court did not accept until 1936.[13] These funds have also allowed Congress effectively to bribe states into exercising their broader police powers as Congress sees fit.[14] Once states are "hooked" on receiving federal funds, they can be coerced to obey federal dictates or lose the revenue.

That the Sixteenth Amendment was necessary to empower Congress to tax incomes is contested. Some maintain that the amendment was only needed to correct an erroneous Supreme Court decision that denied Congress this power.[15] Whatever the merits of this claim, prior to the Sixteenth Amendment, Congress had not taxed income except in times of war.[16] Since 1913, Congress has taxed income at an increasing rate and used the revenues to vastly expand its reach beyond its enumerated powers as even the post-New Deal Supreme Court defines them, co-opting state governments to do its bidding in ways it could not do itself.

The second structural change was the Seventeenth Amendment, providing for the direct election of United States senators by the voters of each state.[17] Under the original Constitution, senators were selected by state legislatures.[18] Senators could therefore be expected to provide some check on the growth of federal power at the expense of the reserved powers of the states. How much of a check on federal interference with state governments this constraint ever provided cannot be assessed with any precision. In addition, the selection of senators by state legislatures was being phased out by the procedure of appointing senators who had prevailed in state elections.[19] Regardless of how effective the previous system may have

---

13. *See* United States v. Butler, 297 U.S. 1, 66 (1936) ("[T]he power of Congress to authorize expenditure of public moneys for public purposes is not limited by the direct grants of legislative power found in the Constitution.").

14. *See* South Dakota v. Dole, 483 U.S. 203, 206 (1987) (allowing federal funding to be conditioned on states exercising their legislative powers as Congress wishes).

15. *See* Pollock v. Farmers' Loan & Trust Co., 157 U.S. 429 (1895) (declaring the Income Tax of 1894 unconstitutional as it violated the requirement that direct taxes be apportioned); *see, e.g.*, MAGLIOCCA, *supra* note 11, at 77 ("Almost nobody prior to *Pollock* thought that Congress lacked the authority to impose an income tax."); *id.* at 76–87 (discussing *Pollock* extensively).

16. *See e.g.*, Revenue Act of 1861, ch. 45, § 49, 12 Stat. 292, 309, *repealed by* Revenue Act of 1862, ch. 119, § 89, 12 Stat. 432, 473. The Revenue Act of 1862 replaced the flat tax imposed by the previous Act with a progressive tax rate. *See id.* § 90, 12 Stat. 473. The Revenue Act of 1862 specified that the income tax it created would terminate after June 30, 1866. *See id.* § 92, 12 Stat. 474.

17. *See* U.S. CONST. amend. XVII.

18. *See* U.S. CONST. art. I, § 3, cl. 1.

19. *See* Senate Historical Office, *Direct Election of Senators*, U.S. SENATE, http://www.senate.gov/artandhistory/history/common/briefing/Direct_Election_Senators.htm (last visited May 5, 2011) (noting that as many as twenty-nine states directly elected senators by 1912).

been, the Seventeenth Amendment eliminated the only structural "check" on federal power that the Founders provided to state governments in the original Constitution.

All this was accomplished because, throughout the twentieth century, the growth of federal power was popular enough that political movements were able to successfully push for constitutional amendments. These groups were also able to elect presidents who would nominate judges who adopted a latitudinarian construction of federal power, along with senators who would confirm them. To the extent that the popularity of unfettered federal power is waning—and only time will tell if this is a blip or a trend—those who would limit federal power need to address the twentieth-century changes to the text of the Constitution and to its interpretation by courts.

A constitutional amendment, or amendments, to constrain Congress is one option. But what sort of amendment? The most obvious type would be a provision like the First Amendment commanding that "Congress shall or shall not do X." However, any additional text that relies on judicial enforcement would likely be undermined by the same post-New Deal judicial philosophy that construed existing constitutional constraints out of existence. What is needed is a structural check on federal power residing not in the judiciary but elsewhere.

One proposal is the Repeal Amendment, which has already been introduced into Congress and will also be considered by state legislatures. The Repeal Amendment would give two-thirds of the states the power to repeal any federal law or regulation. Its text is simple, and its effect is transparent:

> Any provision of law or regulation of the United States may be repealed by the several states, and such repeal shall be effective when the legislatures of two-thirds of the several states approve resolutions for this purpose that particularly describe the same provision or provisions of law or regulation to be repealed.[20]

On May 12, 2011, this proposal was reintroduced before the House of Representatives by Rep. Bob Bishop and introduced before the Senate by Sen. Michael Enzi.[21] In the same month, Florida became the first state legislature to call for an Article V convention to adopt this proposed Amendment.[22]

---

20. H.R.J. Res. 542, 111th Cong. (2010).

21. *See* S.J. Res. 12, 112th Cong. (2011); H.R.J. Res. 62, 112th Cong. (2011); *see generally* Mike Enzi & Bob Bishop, *Introducing the Repeal Amendment*, FOXNEWS.COM (May 12, 2011), http://www.foxnews.com/opinion/2011/05/12/sen-mike-enzi-rep-rob-bishop-introducing-repeal-amendment/.

22. *See* S. Con. Res. 1558, 2011 S., Reg. Sess. (Fla. 2011); H. Mem. 1429, 2011 H., Reg. Sess. (Fla. 2011).

At present, the only way for states to contest a federal law or regulation is to bring a constitutional challenge in federal court or seek an amendment to the Constitution. A state repeal power would provide a targeted way to reverse particular congressional acts and administrative regulations without either relying on federal judges or permanently amending the text of the Constitution just to correct a specific abuse of federal power.

A state repeal power should not be confused with the power of federal courts to "nullify" unconstitutional laws. Unlike the judiciary, under the Repeal Amendment, states can reject a federal law for policy reasons that are irrelevant to constitutional concerns. In this sense, a state repeal power is more like the President's veto power, though it can be applied to *any* existing law or regulation that has already been enacted.

This provision would help restore the original balance between state and federal power and allow states to protect the liberties and rights of their citizens, as well as their own operations, from overreaching federal power. It places confidence in the collective wisdom of the men and women from diverse backgrounds, elected by diverse constituencies, who comprise the modern legislatures of two-thirds of the states. Put another way, it allows thousands of democratically elected representatives outside the Beltway to check the will of 535 elected representatives in Washington, D.C.

Compare this with the presidential veto power held by a single person or the power of five justices to nullify a law they find unconstitutional. But unlike a law declared unconstitutional, nothing in the Amendment would prevent Congress from reenacting a repealed measure if it felt that two-thirds of state legislatures were somehow out of touch with popular sentiment. Unlike the presidential veto, congressional reenactment would require just a simple majority. In effect, with this power the states could force Congress to take a second look at a controversial law.

Americans revere their Constitution, but they have also acted politically to improve it. The Thirteenth[23] and Fourteenth[24] Amendments limited the original power of states to violate the fundamental rights of their own citizens, while the Fifteenth[25] and Nineteenth[26] Amendments prohibited disfranchisement based on race and sex, respectively. Additionally, the Twenty-first Amendment[27] repealed another "progressive" reform: the Eighteenth Amendment that empowered Congress to prohibit alcohol.[28]

The Repeal Amendment alone will not cure all the current problems with federal power. Getting two-thirds of state legislatures to agree on repealing a federal law will not be easy, and repeal will only happen if a law is highly unpopular. Perhaps its most important effect will be deterring

23. *See* U.S. CONST. amend. XIII.
24. *See* U.S. CONST. amend. XIV.
25. *See* U.S. CONST. amend. XV.
26. *See* U.S. CONST. amend. XIX.
27. *See* U.S. CONST. amend. XXI.
28. *See* U.S. CONST. amend. XVIII (repealed 1933).

even *further* expansions of federal power. Just as Congress must now contemplate the President's veto, so too would it need to anticipate how states will react.

While it is no panacea, the Repeal Amendment would restore the states' ability to protect the powers "reserved to the states" noted in the Tenth Amendment.[29] Moreover, it would provide citizens with another political avenue to protect the "rights . . . retained by the people" to which the Ninth Amendment refers.[30] In short, the Repeal Amendment would provide a new political check on the threat to American liberties posed by a runaway federal government.

The Repeal Amendment has already drawn some criticism. First, the *Washington Post*'s Dana Milbank tried to associate the measure with racism: "[T]here's the unfortunate echo of nullification—the right asserted by states to ignore federal laws they found objectionable—and the 'states' rights' argument that was used to justify slavery and segregation."[31] But, this association is imaginary. Undermining civil rights is simply not on the agenda of anyone who favors this amendment. Even before the Civil War, two-thirds of the states never supported slavery or segregation. Had the Repeal Amendment existed then, at least half the states, though not two-thirds, would have used this power in an attempt to repeal the Fugitive Slave Acts, both of which were enacted by Congress.[32] Today, reaching the two-thirds threshold would require the support of many states from different parts of the country, blue as well as red. As will be further explained below, the two-thirds threshold ensures a broad and bipartisan political consensus.

In addition to Milbank's criticism, *Slate*'s Dahlia Lithwick and Jeff Shesol attempted to find a contradiction in Repeal Amendment supporters' professed love for the Constitution: "For a party (whether of the Tea or Grand Old variety) that sees the Constitution as something so perfect as to have been divinely inspired, the idea that it needs to be altered fundamentally is beyond crediting . . . ."[33] But the amendment process of

---

29. *See* U.S. CONST. amend. X.

30. U.S. CONST. amend. IX.

31. Dana Milbank, *A Strange Way to Honor the Founding Fathers*, WASH. POST, Dec. 1, 2010, http://www.washingtonpost.com/wp-dyn/content/article/2010/12/01/AR201012 0105576.html.

32. Fugitive Slave Act of 1850, ch. 60, §§ 1–10, 9 Stat. 462, 462–65 (repealed 1865); Fugitive Slave Act of 1793, ch. 7, §§ 1–4, 1 Stat. 302, 302–05 (repealed 1865); *see* Map of Free and Slave States, SLAVERY, http://www.sonofthesouth.net/slavery/slave-maps/map-free-slave-states.htm (last visited Feb. 16, 2011). In 1857, there were 31 states. Fifteen permitted slavery, and sixteen had either abolished or never permitted slavery. Twenty-one states would have been required to meet the two-thirds threshold.

33. Dahlia Lithwick & Jeff Shesol, *Repealing Common Sense: The Conservative Mission to Destroy the Constitution in Order to Save It*, SLATE, Dec. 3, 2010, http://www.slate.com/id/2276463.

Article V is part of the original Constitution. And, as was noted above,[34] this process has already been used to alter the original scheme by allowing a national income tax and eliminating the power of state legislatures to select United States senators. Add to this that the judicial construction of the Constitution has vastly expanded federal power in just the past sixty years. To the extent Lithwick and Shesol sincerely care about the original Constitution, the Repeal Amendment is simply restoring some semblance of the original state-federal balance.

Then the editors of the *New York Times* weighed in with objections that are a little harder to fathom. They noted, "Under the Tea Party proposal, the states would have much greater power than the president to veto federal laws. Because the amendment includes no limit on the time in which states could exercise their veto, it would cast a long shadow over any program under federal law."[35] Getting both houses of the legislatures of two-thirds of the states to repeal a federal law, however, would be a daunting task—far more difficult than a single President wielding a veto pen. True, older laws can be repealed, but this is likely to happen only when these laws are no longer perceived as current. And inserting states into the actual lawmaking of Congress, as the veto power inserts the President in the legislative process, would raise practical difficulties of its own, so the Repeal Amendment avoids this by operating only after the fact.

The *New York Times* also made a more fundamental objection. It rejected the notion "that the United States defined in the Constitution are a set of decentralized sovereignties where personal responsibility, private property and a laissez-faire economy should reign."[36] Instead, it contended, "America's fundamental law holds competing elements, some constraining the national government, others energizing it."[37] But giving two-thirds of state legislatures a formal way to "constrain[] the national government" no more elevates states into "sovereignties" than the veto power makes the President a king. Giving states this option simply compensates for other changes that have greatly expanded federal power at the expense of the reserved powers of states and the rights retained by the people.

And, while the *New York Times* is quite correct in saying that "the government the Constitution shaped was founded to create a sum greater than the parts, to promote economic development that would lift the fortunes of the American people,"[38] one crucial mechanism by which this was accomplished was through the scheme of checks and balances. A defect in this original scheme was an inadequate federal check on state powers, when such powers were used by states to oppress their own

34. *See supra* notes 12, 17 and accompanying text.

35. Editorial, *The Repeal Amendment*, N.Y. TIMES, Dec. 27, 2010, at A14, *available at* http://www.nytimes.com/2010/12/27/opinion/27mon2.html.

36. *Id.*

37. *Id.*

38. *Id.*

citizens. This defect was rectified by an amendment devised by Republicans in the Thirty-Ninth Congress: The Fourteenth Amendment.[39] What is now lacking is any complementary check on federal power by the states.

In part because the judiciary has failed to exercise its own checking function, the powers of Congress have grown so enormously that they swamp the operations of state governments. For this reason the Court has recognized certain limits on Congress *vis-à-vis* state legislatures[40] and state executive officials.[41] The Repeal Amendment merely places an additional structural check in the hands of democratically elected members of state legislatures.

The only real objection of substance to the Repeal Amendment concerns the theoretical possibility that two-thirds of the least populous states—representing less than half of the nation's population—could stymie legislation backed by a majority. As Milbank observes, "the 33 smallest states, which have 33 percent of the population, have the power to overrule the 17 largest states, which have 67 percent of the population."[42]

Of course, our Constitution is as much about protecting the minority from the tyranny of the majority as it is about majoritarian rule. Indeed, the legitimacy of majority rule is suspect unless it is somehow constrained to protect the rights of individuals from abuses by majorities. In other contexts, one expects that Milbank would agree. As long as one is fantasizing, why is it proper that densely populated, urban states could expropriate the wealth of less populated areas—perhaps to pay their public-sector union workers large pensions? Abuses of creditors, who constituted a small minority of the citizenry, by state legislatures appealing to the large majority of voters who were debtors, was just one of the reasons the Constitution was adopted in the first place.[43] It was certainly not adopted to allow a majority of voters at the national level to exploit, economically or otherwise, a minority, which is the reason why each state is entitled to two senators regardless of population—senators formerly selected by state legislatures.

Nonetheless, any such counter-majoritarian scenario is highly unlikely. Remove just seven of the least populous blue states,[44] add Florida and

---

39. *See* U.S. CONST. amend. XIV, § 5 (giving Congress the power to enforce the Fourteenth Amendment's provisions against the states).

40. *See* New York v. United States, 505 U.S. 144 (1992) (barring federal commandeering of state legislatures).

41. *See* Printz v. United States, 521 U.S. 898 (1997) (barring federal commandeering of state executives).

42. Milbank, *supra* note 31.

43. *See* U.S. CONST art. I, § 10 ("No State shall . . . pass any . . . Law impairing the Obligation of Contracts.").

44. From least to most populous: Vermont, Delaware, Rhode Island, Hawaii, New Hampshire, New Mexico, and Connecticut. U.S. DEP'T OF COMMERCE, U.S. CENSUS

Texas to the thirty-two remaining least populous states,[45] and the result is well over one half of the national population.[46] Furthermore, this group is comprised of a mix of states, both politically and geographically. Given that the political valence of states is not necessarily correlated with their size, the fear of a small state takeover is reminiscent of some of the more fevered writings of the Antifederalists.[47]

Realistically, we can expect two-thirds of state legislatures to band together to repeal a law or regulation under two circumstances. The first is when public opinion turns against a formerly popular law or when, for some unusual reason, a majority of the 535 individuals comprising Congress plus the President become grossly out of step with public opinion. Allowing elected legislators outside the Beltway to check this power is a way of protecting, rather than undermining, truly popular governance.

In the second circumstance, Congress or a regulatory agency may have messed with the internal operation of state governments in ways that are out of public view. Perhaps the regulation is buried in a massive omnibus bill. To claim a majoritarian imprimatur for such a law, or many an administrative regulation, is pure fiction. Such measures have never been subjected to any meaningful popular approval. Of course, if two-thirds of the states take exception and gain its repeal and repeal is unpopular—something pretty hard to imagine—Congress can always then reenact such a measure by a simple majority vote, ensuring that it truly reflects the views of a congressional majority.

This highlights the ultimate safety valve built into the Repeal Amendment: Congress can reenact anything the states manage to repeal. Unlike a constitutional decision of the Supreme Court, Congress could override a state repeal. And it would see a repeal movement coming from a mile away, preparing it to reenact legislation should it feel strongly that the states are misguided or out of touch. At the end of the day, all states may do is force Congress to take a second look at a measure.

BUREAU: Table 2. Resident Population of the 50 states, the District of Columbia, and Puerto Rico: 2010 Census, *available at* http://2010.census.gov/news/press-kits/apportionment/apport.html.

45. Alabama, Alaska, Arizona, Arkansas, Colorado, Idaho, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, North Dakota, Oklahoma, Oregon, South Carolina, South Dakota, Tennessee, Utah, Virginia, Washington, West Virginia, Wisconsin, and Wyoming. *Id.*

46. The result is approximately 166 million. *Id.*

47. *See, e.g.*, Centinel, *Number I (October 5, 1797), in* THE ANTI-FEDERALIST PAPERS AND THE CONSTITUTIONAL CONVENTION DEBATES 235 (Ralph Ketcham ed., Signet Classic 2003) (1986). Centinel argues against bicameralism because the "smallest State in the Union has equal weight with the great States of Virginia, Massachusetts or Pennsylvania." Centinel also stated that the Constitution was "a most daring attempt to establish a despotic aristocracy among freemen, that the world has ever witnessed." *Id.* at 232.

So the strongest objection to the Repeal Amendment may well be that it is too modest to check runaway federal power effectively. But, like the President's veto power, the threat of repeal would deter Congress from interfering with states. Members of Congress and their staffs would have to think about the possible reaction of state legislatures. State legislative leaders could organize to communicate their views to Washington, and the public would have an alternate channel of protest when the federal government gets too out of touch.

That so modest a measure as the Repeal Amendment would so frighten folks like Dana Milbank, Dahlia Lithwick, and the editors of the *New York Times* is a sign of how far federal power has expanded. That these are the strongest objections they could muster shows the strength of the proposal. But the tenor of their reaction also suggests that, however modest it may be, the Repeal Amendment is a genuine step in the right direction.

AMENDING ARTICLE V TO MAKE THE CONSTITUTIONAL AMENDMENT PROCESS ITSELF LESS ONEROUS

TIMOTHY LYNCH*

The subject of this symposium is original, important, and timely.[1] In my view, too much energy in the legal community has been devoted to determining whether a Supreme Court ruling was correct or not. Such work is necessary and helpful, of course, but we should stand back from current events more often and ask the basic questions that are the subject of this symposium. Among these questions is: How can our fundamental legal charter be improved?

To be sure, some deficiencies in the constitutional design were inevitable. Thus, the real question is whether such deficiencies are so bad that they require amendments to our Constitution. I can think of several amendments that would benefit our polity. Many of the amendments that I would support—such as making it more difficult for the American military to go to war—can be fairly characterized as an attempt to "restore" the original understanding of the Constitution.[2] However, for purposes of this symposium, I want to propose a change to the Constitution as it was understood in 1787. My thesis is that the procedure for amending the Constitution in Article V is defective and should be changed. I am, in short, calling for amending the amendment process itself.[3]

I. THE ARTICLE V FRAMEWORK

The framers of the Constitution knew full well that they were incapable of creating a perfect legal charter.[4] They appreciated the fact that times change and that it may be necessary and desirable to amend the Constitution as various problems and situations arose.[5] Under the Articles

---

* Timothy Lynch is the director of the Cato Institute's Project on Criminal Justice.

1. *See, e.g.*, Robert Barnes, *Scholars Debate Whether Time is Right for Amending the Constitution*, WASH. POST (Nov. 28, 2010, 6:43 PM), http://www.washingtonpost.com/wp-dyn/content/article/2010/11/28/AR2010112803275.html.

2. *See* GENE HEALY, THE CULT OF THE PRESIDENCY 30–32 (2008).

3. Other scholars have reached the same conclusion. *See, e.g.*, Stephen M. Griffin, *The Nominee is . . . Article V*, 12 CONST. COMMENT. 171 (1995); Michael B. Rappaport, *Reforming Article V: The Problems Created by the National Convention Method & How to Fix Them*, 96 VA. L. REV. 1509 (2010).

4. This is demonstrated by the mere fact that the framers included Article V in the Constitution, which outlines the manner in which to amend the Constitution. *See* U.S. CONST. art. V.

5. *See* Douglas Linder, *What in the Constitution Cannot be Amended?*, 23 ARIZ. L.

of Confederation, all thirteen state legislatures had to approve any proposed amendment.[6] Many have viewed that high threshold as a terrible defect because badly needed reforms too often languished.[7] The amendment process created by the framers in the new Constitution relaxed the unanimity requirement but kept the bar high by requiring a supermajority among the states.[8] The procedure that they devised is set forth in Article V:

> The Congress, whenever two thirds of both Houses shall deem it necessary, shall propose Amendments to this Constitution, or, on the Application of the Legislatures of two thirds of the several States, shall call a Convention for proposing Amendments, which in either Case, shall be valid to all Intents and Purposes, as Part of this Constitution, when ratified by the Legislatures of three-fourths of the several States, or by Conventions in three fourths thereof, as the one or the other Mode of Ratification may be proposed by the Congress; . . . [although] no State, without its Consent, shall be deprived of its equal Suffrage in the Senate.[9]

In *The Federalist No. 43*, James Madison made the case that Article V strikes the right balance between two possible errors: "It guards equally against that extreme facility, which would render the Constitution too mutable; and that extreme difficulty, which might perpetuate its discovered faults."[10]

Over the past 224 years, the Constitution has been amended seventeen times.[11] There have been twenty-seven amendments, but the first ten amendments—the Bill of Rights—were a package proposed by the very first Congress.[12] There has never been a successful call for a constitutional convention.[13] However, the movement favoring the direct election of senators did come close.[14] Congress proposed the Seventeenth Amendment when it became clear that a sufficient number of states would call a convention to enact that reform.[15]

REV. 717, 719 (1981).

6. *See* ARTICLES OF CONFEDERATION of 1788, art. XIII.

7. *See, e.g.*, THE FEDERALIST NO. 43, at 245 (James Madison) (E.H. Scott ed., 1898).

8. *See* U.S. CONST. art. V.

9. *Id.*

10. THE FEDERALIST NO. 43, at 245 (James Madison) (E.H. Scott ed., 1898).

11. *See* Griffin, *supra* note 3, at 172.

12. *See* Stephen M. Griffin, *The Problem of Constitutional Change*, 70 TUL. L. REV. 2121, 2135 (1996).

13. *See id.*

14. *See* Gerard N. Magliocca, *State Calls for an Article Five Convention: Mobilization and Interpretation*, 2009 CARDOZO L. REV. DE NOVO 74, 79–81 (2009).

15. *See id.*

II. THE PROBLEM

Article V was criticized from the start. During the Virginia ratification debates, Patrick Henry argued that the supermajority requirement for amendments was too high a threshold:

> To suppose that so large a number as three-fourths of the States will concur, is to suppose that they will possess genius, intelligence, and integrity, approaching to miraculous. It would indeed be miraculous that they should concur in the same amendments, or, even in such as would bear some likeness to one another. For four of the smallest States, that do not collectively contain one-tenth part of the population of the United States, may obstruct the most salutary and necessary amendments . . . .[16]

According to Henry, "the way of amendment" was, effectively, "shut."[17] The early experience under the new Constitution seemed to dispel Henry's fears. As noted, the first Congress proposed the Bill of Rights, and those safeguards were promptly ratified.[18] Thus, the amendment procedure appeared to work just fine. And as the years passed, Americans became accustomed to the amendment process and the new constitutional regime in general.

Nowadays, journalists, historians, and others express awe at the relative paucity of amendments to the American Constitution.[19] They marvel at how a charter drafted in 1787 could bring America into the twenty-first century with so few changes.[20] This paucity is offered as evidence of the genius of the Constitution's overall design. But that claim is wrong—and profoundly so. The truth of the matter is that the original understanding of the Constitution has eroded over time. Disheartened by the chances of successfully amending the Constitution, political activists, reformers, and politicians began embracing a strategy of accomplishing their objectives by outwardly voicing respect for the Constitution while working assiduously for a "reinterpretation" of the document to allow for the laws and powers that they deemed beneficial to the country.[21]

This is not the place for a wide-ranging examination of the erosion, but the extraordinary interpretation that academics, lawyers, elected officials, and jurists have given to Congress's power "[t]o regulate commerce" should suffice for present purposes.[22] To begin with, the Constitution

---

16. *Speeches of Patrick Henry (June 5 and 7, 1788)*, *in* THE ANTI-FEDERALIST PAPERS AND THE CONSTITUTIONAL CONVENTION DEBATES 204 (Ralph Ketcham ed., 1986).

17. *Id.* at 203.

18. *See* Griffin, *supra* note 12, at 2135.

19. *See id.* at 2122–23; Magliocca, *supra* note 14, at 74–76.

20. *See* Griffin, *supra* note 12, at 2122–23.

21. *See generally* RICHARD A. EPSTEIN, HOW PROGRESSIVES REWROTE THE CONSTITUTION (2006).

22. U.S. CONST. art. 1, § 8.

creates a federal government of limited and enumerated powers.[23] In *The Federalist No. 45*, Madison observed, "The powers delegated by the proposed Constitution to the Federal Government, are few and defined. Those which are to remain in the State Governments, are numerous and indefinite."[24] Most of the federal government's "delegated powers" are specified in Article I, § 8.[25] The Tenth Amendment was appended to the Constitution to make it clear that the powers not delegated to the federal government are "reserved to the States respectively, or to the people."[26]

Today, the federal government consists of dozens of regulatory agencies and spends trillions of dollars—activities that would have been unimaginable to the people who debated and ratified the Constitution.[27] Few people claim that the Constitution is inadequate.[28] Instead, others have claimed that the language in Article I, § 8 was phrased in such broad terms as to allow for the expansion of federal power.[29]

Consider the landmark case *United States v. Lopez*.[30] The facts in the *Lopez* case are straightforward. Alphonso Lopez was a high school student who was caught carrying a handgun on school premises.[31] He was initially arrested under a Texas law that prohibited the possession of firearms on school property, but federal agents took over the case and prosecuted Lopez for violating the "Gun-Free School Zones Act of 1990."[32] After his conviction, Lopez's attorney argued on appeal that Congress had exceeded its constitutional authority when it passed the law.[33] When a federal appeals court agreed with Lopez and overturned his conviction, the Justice

---

23. *See* U.S. CONST. amend. X (limiting the federal government's powers to those "delegated to the United States by the Constitution").

24. THE FEDERALIST NO. 45, at 258 (James Madison) (E.H. Scott ed., 1898).

25. *See* U.S. CONST. art. 1, § 8.

26. U.S. CONST. amend. X.

27. *See, e.g.*, RANDY E. BARNETT, RESTORING THE LOST CONSTITUTION 316–17 (2004).

28. There are a few exceptions. *See generally* SANFORD LEVINSON, OUR UNDEMOCRATIC CONSTITUTION (2006) (assessing the defects of the Constitution); LARRY J. SABATO, A MORE PERFECT CONSTITUTION (2007) (suggesting twenty-three proposals to revitalize the Constitution).

29. Interestingly, Rexford Tugwell, a "principal member of President [Franklin] Roosevelt's 'Brain Trust,'" admitted that the Supreme Court rulings upholding the constitutionality of the New Deal programs "were 'tortured interpretations of a document intended to prevent them.'" Roger Pilon, *Freedom, Responsibility, and the Constitution: On Recovering Our Founding Principles*, 68 NOTRE DAME L. REV. 507, 516 n.36 (1993) (quoting Rexford G. Tugwell, *Rewriting the Constitution: A Center Report*, CENTER MAG., Mar. 1968, at 20).

30. United States v. Lopez, 514 U.S. 549 (1995); *see also* Richard A. Epstein, *Constitutional Faith and the Commerce Clause*, 71 NOTRE DAME L. REV. 167 (1996) (discussing *Lopez* and its potential impact on the Commerce Clause).

31. *Lopez*, 541 U.S. at 551.

32. *Id.*

33. *See id.* at 552.

Department appealed the case to the Supreme Court, which agreed to hear the case.[34]

In his brief to the Supreme Court, Solicitor General Drew Days argued that the Gun-Free School Zones Act could be justified under Congress's power "to regulate Commerce with foreign Nations, and among the several States, and with the Indian Tribes."[35] Days argued that possession of a gun in a school zone: (a) might lead to violent crime, which (b) might threaten the learning process, which (c) might ultimately produce less productive citizens, which (d) might, cumulatively, impair the national economy and interstate commerce.[36]

When Days appeared before the Court for oral argument, the justices pressed him on the implications of his constitutional theory:

> QUESTION: General Days, just to understand what we're talking about, do I correctly understand your position to be, your rationale for this--
>
> GENERAL DAYS: Yes.
>
> QUESTION: --that all violent crime, if Congress so desired, could be placed under a Federal wing, could be placed in the Federal court for prosecution, all violent crime, or is there any stopping point? Is there any violent crime that doesn't affect interstate commerce on you[r] rationale?
>
> GENERAL DAYS: Well, Your Honor, I think the answer is that it may be possible for Congress to do that under the commerce power. . . .
>
> QUESTION: [So] there is no question that Congress has the power, in effect, to take over crime, because I--
>
> GENERAL DAYS: I do not--
>
> QUESTION: --presume there's no limitation on your rationale, or on Congress' rationale, that would preclude it from reaching any traditional criminal activity.
>
> GENERAL DAYS: That's correct.[37]

---

34. *See id.*

35. *See* Brief for the United States at 2–6, United States v. Lopez, 514 U.S. 549 (1995) (No. 93-1260) [hereinafter *Lopez* Brief].

36. *See id.* at 19–25.

37. Transcript of Oral Argument at 10–13, United States v. Lopez, 514 U.S. 549 (1995) (No. 93-1260) [hereinafter *Lopez* Oral Argument]; *see also* Lyle Denniston, *Going Overboard for a Federal Law*, 17 AM. LAWYER 94–95 (1995) (describing Solicitor General Days's "daring claim of power for Congress").

As that exchange makes clear, the stakes in *Lopez* went well beyond the constitutionality of the Gun-Free School Zones Act.[38] Attorneys for the federal government outlined a radically expansive theory of federal power.[39] Solicitor General Days maintained that the federal government could not only fight all types of violent crime, but could regulate any activity that might lead to violent crime.[40] Days also argued that he could discern no limit on Congress's power to regulate commerce—it was, for all intents and purposes, "plenary."[41]

The Supreme Court recoiled from the federal government's position: "[i]f we were to accept [Days's] arguments, we are hard pressed to posit any activity by an individual that Congress is without power to regulate."[42] In his concurring opinion, Justice Clarence Thomas noted that, if Congress had been given authority over matters that simply "affect" interstate commerce, much, if not all, of the enumerated powers set forth in Article I, § 8 would be unnecessary.[43] Indeed, it is difficult to dispute Justice Thomas's conclusion that an interpretation of the commerce power that "makes the rest of § 8 superfluous simply cannot be correct."[44]

Much of what the federal government does today is inconsistent with what is supposed to be the supreme law of the land.[45] The danger was acknowledged early on: Thomas Jefferson wrote, "Our peculiar security is in the possession of a written Constitution. Let us not make it a blank paper by construction."[46] Unfortunately, the original understanding of the

---

38. *See Lopez* Oral Argument, *supra* note 37, at 10–13.

39. *See generally Lopez* Brief, *supra* note 35 (arguing that the United States used Article 1, Section 8, Clauses 3 and 18 of the Constitution to demonstrate that Congress has the power to legislate under the Commerce Clause.).

40. *See id.* at 18 n.11.

41. In oral argument before the Supreme Court, Solicitor General Days stated, "Congress was legislating for the entire Nation pursuant to its plenary powers under the Constitution." *Lopez* Oral Argument, *supra* note 37, at 4.

42. United States v. Lopez, 514 U.S. 549, 564 (1995).

43. *See id.* at 589 (Thomas, J., concurring).

44. *Id.* (Thomas, J., concurring).

45. *See generally* Gary Lawson, *The Rise and Rise of the Administrative State*, 107 HARV. L. REV. 1231 (1994) (detailing the history of the administrative state and arguing that it is unconstitutional). One other example that is worth a brief mention concerns the modern "war on drugs." *See* Gonzalez v. Raich, 545 U.S. 1, 10 (2005). Note that the Constitution was amended to permit the federal government to police a ban on the manufacture of liquor. *See* Roger Pilon, *The Illegitimate War on Drugs*, *in* AFTER PROHIBITION: AN ADULT APPROACH TO DRUG POLICIES IN THE 21ST CENTURY 23, 26–27 (Timothy Lynch ed., 2000). Instead of seeking an amendment to permit federal agents to police a ban on narcotics, the Commerce Clause was simply stretched to accommodate the policy. *See id.* (discussing the history of the Commerce Clause and its relation to the war on drugs); *see also Gonzales*, 545 U.S. at 57 (Thomas, J., dissenting) (discussing the authority and the limits of the Commerce Clause).

46. Letter from Thomas Jefferson to Wilson C. Nicholas (Sept. 7, 1803), *in* 10 THE

Constitution has been corrupted by construction. By making the Constitution difficult to amend, the framers thought they could preserve it, but the design failed. It seems plain that our courts and policymakers are not seriously committed to the original understanding of the charter.[47]

III. AMENDING ARTICLE V

In the previous section, I showed that the Constitution of 1787 is not working as it was designed to work. In many key aspects, this charter has been reinterpreted to allow the federal government to expand beyond the limited powers in the original design. This is not a new development. We have been on this course for some time and can continue to muddle along in this fashion. In my view, however, this is a profoundly unsettling state of affairs. In light of the erosion in the original design of the Constitution, how can anyone be confident that other constitutional safeguards will not be lost?

I also maintain that the difficult amendment procedure laid out in Article V is *primarily* responsible for our current predicament. Admittedly, that is hard to prove, but the political scientist Donald Lutz has studied constitutional charters from around the world and has shown that the American Constitution is among the most difficult to amend.[48]

Given the extraordinary growth of the federal government, one would have to heavily discount any claim that the low rate of amendment has been due to a general satisfaction with the Constitution of 1787. A more plausible explanation, I submit, is that the fervor for change in the role of government has followed the path of least resistance, which has essentially meant finding like-minded jurists, fighting over nominations to the Supreme Court, and mounting legal defenses when federal powers have been challenged in the courts.[49]

WRITINGS OF THOMAS JEFFERSON 417, 419 (Andrew A. Lipscomb & Albert E. Bergh eds., 1903).

47. *See generally* Gary Lawson, *Making a Federal Case Out of It:* Sabri v. United States *and the Constitution of Leviathan*, *in* CATO SUPREME COURT REV. 119 (2004) (explaining how the Supreme Court used faulty constitutional principles in deciding *Sabri v. United States*); Timothy Lynch, *A Smooth Transition: Crime, Federalism, and the GOP*, *in* THE REPUBLICAN REVOLUTION 10 YEARS LATER 213 (Chris Edwards & John Samples eds., 2005) (discussing the Republicans' treatment of the Constitution while in power from 1994 through 2004); Ilya Somin, *Taking Stock of* Comstock*: The Necessary and Proper Clause and the Limits of Federal Power*, *in* CATO SUPREME COURT REVIEW 239 (2010) (discussing how the Supreme Court interpreted the Constitution in *United States v. Comstock*).

48. According to Lutz, the United States has one of the lowest rates of amendment worldwide. *See* Donald Lutz, *Toward a Theory of Constitutional Amendment*, 88 AM. POL. SCI. REV. 355 (1994).

49. *See generally* Roger Pilon, *How Constitutional Corruption Has Led to Ideological Litmus Tests for Judicial Nominees*, 446 CATO INST. POLICY ANALYSIS (Aug. 6, 2002) (portraying how Americans are more concerned that the Supreme Court justices' views are

If Article V is indeed primarily responsible for the problem, the next question concerns the remedy. While I am firmly convinced on the need for some relaxation in the threshold necessary to secure the adoption of an amendment, I am less committed to any particular proposal that would accomplish that end. With that caveat, let me advance a specific proposal in the hope of generating more thought and discussion in this direction.

First, Article V divides the amendment procedure into two distinct phases.[50] Phase one initiates the amendment process, and phase two concerns the ratification process.[51] This division seems prudent and ought to be retained. But instead of the two-thirds vote necessary for Congress to propose amendments or for the states to call a convention, we should lower the threshold to a simple majority. And instead of the three-fourths vote necessary for the states to ratify an amendment, we should lower that threshold to two-thirds.

A skeptic might reasonably ask whether my proposed amendment could bring a different set of problems into American politics. Of course it could. First, we could see an uptick in the number of amendments proposed and ratified. Second, many of us might dislike or even deplore some of these amendments. The key question, however, is whether these problems will be worse than the predicament in which we find ourselves in 2011. To that question, my answer is "No." An easier amendment process will bridge the gulf that presently exists between the constitutional text and the government we actually have. Also, an easier amendment process will bring more candor and less cynicism to constitutional discourse. In the long run, I would also expect an easier amendment process to enliven our politics in a way that would be healthy for our republic.

---

consistent with the views of the American people, rather than whether they will apply the law).

50. *See* U.S. CONST. art. V.

51. *See id.*

THE TRUTH-IN-LEGISLATION AMENDMENT: AN IDEA WHOSE TIME HAS COME

BRANNON P. DENNING* & BROOKS R. SMITH**

If a constitutional convention were called tomorrow and the delegates solicited proposals for topics, we would dust off a proposal we made over ten years ago to subject congressional legislation to a single-subject rule similar to those found in most state constitutions.[1] As Congress continues to pass legislation that its members cannot possibly have read or understood beforehand, and then when members of Congress tell the public that they must first pass legislation to find out what is in it,[2] only to discover that the contents are deeply unpopular, we think it is time to renew our call for the "Truth-in-Legislation Amendment." This article will lay out our proposal and our rationale for its passage. We will also anticipate some criticisms that went unaddressed in our earlier article.

I.

The Truth-in-Legislation Amendment (TILA) would be modeled on provisions found in most state constitutions.[3] It would read:

> Section 1. Congress shall pass no bill, and no bill shall become law, which embraces more than one subject, that subject being clearly expressed in the title.
>
> Section 2. Notwithstanding the requirements of Article III, any taxpayer and any member of Congress shall have standing to enforce the provisions of Section 1 by filing suit in federal district court.

As noted in our earlier article, similar provisions began to appear in state constitutions during the mid-nineteenth century.[4] The first single-subject

---

* Professor, Cumberland School of Law, Samford University.

** Partner, Bradley Arant Boult Cummings, LLP, Nashville, Tennessee. The opinions expressed herein are those of the authors and not necessarily those of Bradley Arant Boult Cummings, LLP or its clients.

1. *See generally* Brannon P. Denning & Brooks R. Smith, *Uneasy Riders: The Case for a Truth-in-Legislation Amendment*, 1999 UTAH L. REV. 957 (1999).

2. *E.g.*, Marguerite Higgins, Video of the Week: *"We have to pass the bill so you can find out what is in it,"* THE FOUNDRY (Mar. 10, 2010, 3:30 PM), http://blog.heritage.org/2010/03/10/video-of-the-week-we-have-to-pass-the-bill-so-you-can-find-out-what-is-in-it/ (recapping a speech on health care reform legislation by then-Speaker Nancy Pelosi (D-CA)).

3. *See* Denning & Smith, *supra* note 1, at apps. A & B.

4. *See id.* at 965–67; *see also* Millard H. Ruud, *"No Law Shall Embrace More Than One Subject,"* 42 MINN. L. REV. 389 (1958).

provision, which appeared in the New Jersey constitution in 1844, provides that its purpose is "[t]o avoid improper influences which may result from intermixing in one and the same act such things as have no proper relation to each other."[5]

By confining legislation to a single subject, legislators will be unable to append extraneous or unrelated amendments to must-pass legislation or to trade the appendage of such riders for support of the legislation itself.[6] Moreover, requiring the subject to be expressed in the title forces legislators to establish the purpose of the legislation and to provide a kind of baseline for evaluating the bill's provisions for compliance with the single-subject requirement.[7] In theory, this combination would enhance transparency in the legislative process—something the federal process seems of late to have lacked. Specifically, we hoped that a federal TILA might curb the use of riders and force the break-up of omnibus bills that had come into frequent use.[8] We hoped that the increased transparency would encourage greater accountability among members of Congress by, for example, "requiring legislators to submit their pork to scrutiny through the normal lawmaking process," thereby "clarify[ing] who is really responsible for legislative boondoggles."[9]

We also tied our proposal to what Hans Linde termed "due process of lawmaking."[10] In a nutshell, Linde argues that the means legislatures use in pursuit of their ends are as important to the legitimacy of the lawmaking process as the ends themselves.[11] Indeed, as we wrote in 1999: "At the heart of Linde's due process of lawmaking model . . . is a concern with procedural integrity and legislative honesty, which in turn assure that the substance of the legislative process is seen by the public as legitimate."[12]

Public concerns over the passage of the Patriot Act and health care, banking, and financial services reforms demonstrate that these problems of legislative legitimacy have not disappeared in the last decade.[13] It seems that even the bills' putative authors hardly know what is in them.[14] Surely the perception of legislative chicanery in the passage of landmark legislation, the disclosure of earmarks submerged in appropriations bills,

---

5. N.J. CONST. art. IV, § 7, ¶ 4; *see also* Denning & Smith, *supra* note 1 at app. B.

6. *See* Denning & Smith, *supra* note 1, at 972.

7. *See id.* at 974.

8. *See id.* at 971–74.

9. *Id.* at 974.

10. *See id.* at 981–86; Hans A. Linde, *Due Process of Lawmaking*, 55 NEB. L. REV. 197 (1976).

11. *See* Linde, *supra* note 10, at 200.

12. Denning & Smith, *supra* note 1, at 984.

13. *See* Jason Iuliano, *Eliminating Earmarks: Why the Congressional Line Item Vote Can Succeed Where the Presidential Line Item Veto Failed*, 112 W. VA. L. REV. 947, 957 (2010).

14. *See* Denning & Smith, *supra* note 1, at 959.

and a kind of ends-justify-the-means attitude in the passage of bills have contributed to Congress's abysmal approval ratings.[15]

Public dissatisfaction with Congress is, of course, hardly new.[16] And we are not naïve enough to think that a single proposal like ours could inaugurate a golden age of good government. But we think that Linde was onto something when he suggested that perceptions of legitimacy in the exercise of political power are influenced by the degree to which legislators are perceived to "play by the rules" when considering and passing legislation.[17]

II.

Our earlier article anticipated some objections to the proposal, ranging from arguments that pork and riders are necessary lubricants to our legislative system,[18] to concerns about the efficacy of similar state provisions,[19] to concerns that the TILA would be *too* effective and would empower the judiciary to paralyze Congress.[20] But perhaps the most powerful potential objections is one that we did not address: whether and how the judiciary could effectively enforce the TILA. In other words, can the judiciary create effective legal doctrine that will help ensure optimal enforcement of the TILA?[21] In the remainder of this article, we consider several interpretive issues courts would face while attempting to enforce the TILA. First, can courts develop a workable definition of "subject?" Second, how would courts enforce the TILA in the context of appropriations bills? Finally, what should courts order as a remedy, rescission of the offending provision or invalidation of the legislation as a whole?

A.

The biggest interpretive challenges for a court would come defining "subject" adequately and providing rules for determining whether a provision in a bill—perhaps embedded in a long and complicated bill—is germane enough to the legislation's overall subject to survive judicial review.[22] Courts would also need to specify who bears the burden of proof

---

15. *See* Iuliano, *supra* note 13, at 957.

16. *See* Denning & Smith, *supra* note 1, at 957.

17. *See id.* at 982–83; Linde, *supra* note 10 at 239, 241.

18. *See id.* at 988–92.

19. *See id.* at 993–1000 (addressing arguments that single-subject provisions have not been effective at curbing riders and logrolling).

20. *See id.* at 1001–03 (noting concerns that the provisions are too formalistic for modern governance and would encourage over-enforcement by the judiciary).

21. *See generally* RICHARD H. FALLON, JR., IMPLEMENTING THE CONSTITUTION (2001); Mitchell N. Berman, *Constitutional Decision Rules*, 90 VA. L. REV. 1 (2004).

22. *See* Michael D. Gilbert, *Single Subject Rules and the Legislative Process*, 67 U.

and what quantum of evidence would satisfy the burden to uphold or invalidate the provision.[23]

Perhaps we ought to begin by making the affirmative case that these decisions are within courts' institutional capabilities. When answering this question, we are put in mind of the preacher who, when asked by a member of his congregation whether he believed in infant baptism, replied, "Believe in it? I've actually seen it done!"[24] Some state courts have had more than 150 years of experience implementing similar constitutional provisions.[25] However, as one recent commentator put it:

> The single subject rule remains, even after a century-and-a-half of life, a source of uncertainty. Not all courts recognize all of the purposes of the rule, and among the purposes they do recognize, there is sometimes hesitancy to flesh them out. Resolution of single subject disputes turns on vague tests that rely as much on judicial commonsense as legal analysis.[26]

Despite the shortcomings of existing single-subject jurisprudence, the large body of case law furnishes an exemplar—even if a negative exemplar—that federal courts might profitably study to enforce the TILA.[27] Moreover, recent scholars and commentators have attempted to sketch workable formulae for courts.[28]

For example, in a recent article, Michael Gilbert has employed public choice theory to solve single-subject problems that have bedeviled courts for more than a century.[29] Gilbert argues rather persuasively that "logrolling"—the practice of legislative vote trading and long a celebrated reason for single-subject requirements—is not inherently harmful.[30] It can produce net social gains as well as losses.[31] By contrast, "riders"—those provisions that, but for their attachment to a popular bill, would likely *not* command majority support—involve not exchange (as does logrolling), "but rather manipulation of legislative procedures."[32] Gilbert defines a rider

---

PITT. L. REV. 803, 806–07 (2006).

23. *See id.* at 807.

24. THE PREACHER JOKE BOOK 41 (Loyal Jones ed., 1989).

25. For our own summary of some of those decisions, *see* Denning & Smith, *supra* note 1, at 993–1000.

26. Gilbert, *supra* note 22, at 829.

27. *See* Denning & Smith, *supra* note 1, at 994.

28. *See id.*; Martha J. Dragich, *State Constitutional Restrictions on Legislative Procedure: Rethinking the Analysis of Original Purpose, Single Subject, and Clear Title Challenges*, 38 HARV. J. ON LEGIS. 103 (2001).

29. *See* Gilbert, *supra* note 22; *see also* Robert D. Cooter & Michael D. Gilbert, *A Theory of Direct Democracy and the Single Subject Rule*, 110 COLUM. L. REV. 687 (2010) (articulating rules for single-subject provisions governing initiatives and referenda).

30. *See* Gilbert, *supra* note 22, at 809.

31. *See id.* at 833–36.

32. *Id.* at 837.

as "a political measure that lacks majority support on its merits but whose opponents vote for it in sufficient numbers to ensure its passage despite not receiving compensation from the measure's supporters."[33]

Gilbert argues that courts should permit logrolling because "every instance of it improves the well being of a majority of legislators," presumably because each gains something during the exchange.[34] He argues that courts should, on the other hand, invalidate riders: "When a bill containing a rider is passed, a majority of legislators is left worse off. They oppose the rider on its face and received nothing in exchange for their support of it."[35] In Gilbert's view, "[e]very bill containing a rider should be condemned for violating the single subject rule."[36]

How then should courts distinguish between permissible logrolling and presumptively invalid riders? Again, Gilbert argues that courts have generally failed to produce tests that are neither tautological nor essentially arbitrary because, at some level of abstraction, every item in any given bill could be said to have been part of a single subject.[37] Instead, he urges judges to inquire into "functionally related components" and to identify whether, if a particular component was "removed and voted upon separately," that component would receive majority support.[38] "If the answer is no," Gilbert says, "the component is a rider, and the bill violates the single subject rule."[39] Gilbert further suggests that judges consult the legislative process that produced the bill (whether provisions were added in committee, for example), any legislative history of the bill, and even "voting records, political affiliation, and . . . poll data to hypothesize how legislators would vote on a truncated bill."[40] If courts clearly articulate that this information must be available to support a claim for violation of the single-subject provision, then parties will have incentives to produce it.[41] Gilbert speculates that it might encourage careful recordkeeping and the retention of legislative history in states that do not currently preserve it, so that riders and mere logrolling could be distinguished.[42]

The point is not, of course, that Gilbert's solution to the problem of judicial standards is the best or the only one.[43] Rather, we would contend

---

33. *Id.* (emphasis omitted).

34. *See id.* at 849.

35. *Id.* at 858.

36. *Id.* at 859.

37. *See id.* at 829–30; *see also* Cooter & Gilbert, *supra* note 29, at 710 ("There is no workable theory of interpretation for the single subject rule.").

38. Gilbert, *supra* note 22, at 860–61.

39. *Id.* at 861; *see also* Cooter & Gilbert, *supra* note 29, at 720–21 (discussing the "separable preferences" of voters in applying single-subject provisions to initiatives and referenda).

40. Gilbert, *supra* note 22, at 863 (footnote omitted).

41. *See id.* at 863–64.

42. *See id.*

43. We are not entirely convinced, for example, that logrolling is the benign legislative

that Gilbert's thoughtful attention to the issue, along with existing state court jurisprudence, suggests at least that federal judges applying the TILA would not be required to engage in doctrinal design completely from scratch.

B.

A second important issue is whether to exempt appropriations bills from single-subject provisions, as some states do.[44] As written, the TILA does not allow for this exemption, with good reason: must-pass appropriations bills are often attractive vehicles to which legislators attach riders.[45] Naturally, applying TILA to appropriations bills is hardly a panacea. Earmarked money in appropriations bills, for example, would not violate the TILA as long as the earmarks were related to the subject of the particular appropriations bill (e.g., a transportation bill earmarking money for road projects in particular districts).

C.

One further issue that will confront courts is whether the remedy for violating a single-subject or title provision is to invalidate the entire legislation or simply to sever the offending provision. As Gilbert notes, "severing is an attractive option. When riders are present, they can be excised, and the popular and logrolled provisions of the bill can be left intact. This rewards legislators who enacted legislation through appropriate channels."[46]

Nevertheless, Gilbert rightly points out that severance "fails to provide legislators with an incentive not to engage in this behavior. Indeed, severing encourages legislators to attach riders: with any luck, they will go undetected and become law, and if they are detected, they will simply be removed and can be reattached to another bill."[47] Better to invalidate the entire bill, thereby forcing rider-attaching legislators to internalize the costs of their behavior. Gilbert writes that "[o]ther legislators, whose hard-fought political bargain was undone because of the rider, may be incensed and less likely to bargain with the culprits, and citizens may be enraged by the delay or failure to enact important legislation."[48]

---

phenomenon Gilbert makes it out to be.

44. *See* Gilbert, *supra* note 22, at 824. Even when a constitution does not exempt appropriations bills explicitly, courts sometimes apply more relaxed scrutiny to such bills. *See id.*

45. *See* Denning & Smith, *supra* note 1, at 961.

46. Gilbert, *supra* note 22, at 867.

47. *Id.*

48. *Id.* at 867–68 (footnote omitted).

We find Gilbert's case for invalidation as opposed to severance convincing and would urge courts to adopt it in implementing the TILA. In addition to creating proper incentives for legislators, we also note that the invalidation remedy—and the TILA in general—could also strengthen the President's veto power.[49] Presidents are often forced to accept riders and other provisions in must-pass legislation for fear of a popular backlash stemming from the veto of a bill containing essential or popular provisions.[50]

CONCLUSION

The TILA would not be a cure-all for the legislative (or political) pathologies that plague our system of government. But for all their faults, single-subject and title requirements have served to curb some egregious abuses of the legislative process in the states for over a century and a half. We think that it is time to incorporate the TILA into the U.S. Constitution as both a symbolic reaffirmation of the importance of due process of lawmaking and as a powerful weapon for lawmakers and citizens when Congress falls short of those standards. Obviously we lack the space to canvass all the issues raised by the prospect of enforcing such a provision, but we do hope our short article will at the very least begin a conversation and debate.

49. *See* Denning & Smith, *supra* note 1, at 1000.

50. *See id.* at 972.

AN EQUAL RIGHTS AMENDMENT TO MAKE WOMEN HUMAN

ANN BARTOW*

I can state with some authority that two times fourteen is twenty-eight, flouting the stereotype that women are inept at mathematics and simultaneously framing my argument in favor of an Equal Rights Amendment (ERA). Though the Fourteenth Amendment[1] provides women with partial legal armament (a dull sword, a small shield), equal protection requires something twice as powerful in the form of a Twenty-Eighth Amendment that would expressly vest women with equal rights under the law. The Fourteenth Amendment has completed only half of the job.

Alice Paul, founder of the National Women's Party, first proposed an Equal Rights Amendment in 1923.[2] It was finally passed by Congress in 1972, but at its June 30, 1982, deadline the ERA had been ratified by only thirty-five states, three short of the thirty-eight required for ratification.[3] The entire text of the so far failed Amendment is:

> Section 1. Equality of rights under the law shall not be denied or abridged by the United States or by any state on account of sex.
>
> Section 2. The Congress shall have the power to enforce, by appropriate legislation, the provisions of this article.
>
> Section 3. This amendment shall take effect two years after the date of ratification.[4]

The proposed amendment is surprisingly pithy, given how much pitched opposition it has engendered.[5] Amending the Constitution to make it clear that government actors cannot disadvantage or oppress people based on a characteristic that the Law generally treats as immutable[6] is

---

* Professor of Law, University of South Carolina School of Law. Professor Bartow is joining the faculty of the Pace Law School in May of 2011. She thanks the following people for helpful input with this essay: Anita Bernstein, Bridget Crawford, Susan Kuo, Amy Milligan, Ellen Podgor and Kaimipono Wenger. She thanks Glenn Reynolds for the opportunity to contribute this essay, and dedicates it to Casey Bartow-McKenney and the memory of C. Edwin Baker.

1. U.S. CONST. amend. XIV.

2. *See* Roberta W. Francis, *The History Behind the Equal Rights Amendment*, THE EQUAL RIGHTS AMENDMENT, http://www.equalrightsamendment.org/era.htm (last visited May 5, 2011).

3. *See id.*

4. S.J. Res. 10, 110th Cong. (2007); H.R.J. Res. 40, 110th Cong. (2007).

5. *See* Francis, *supra* note 2.

6. I acknowledge that some people view gender as fluid, and I do not mean to suggest that people cannot change their "official" gender. Nor do I endorse social practices that force

objectionable to people who believe that women as a class need and receive special legal protections linked to sex.[7] Women have sex, both normatively and descriptively, but are women human? Author Dorothy Sayers posed this query in 1938 as the title of a speech, in which she observed:

> A man once asked me—it is true that it was at the end of a very good dinner, and the compliment conveyed may have been due to that circumstance—how I managed in my books to write such natural conversation between men when they were by themselves. Was I, by any chance, a member of a large, mixed family with a lot of male friends? I replied that, on the contrary, I was an only child and had practically never seen or spoken to any men of my own age till I was about twenty-five. "Well," said the man, "I shouldn't have expected a woman [meaning me] to have been able to make it so convincing." I replied that I had coped with this difficult problem by making my men talk, as far as possible, like ordinary human beings. This aspect of the matter seemed to surprise the other speaker; he said no more, but took it away to chew it over. One of these days it may quite likely occur to him that women, as well as men, when left to themselves, talk very much like human beings also.[8]

Sayers asserted that, in her experience, "both men and women are fundamentally human, and that there is very little mystery about either sex, except the exasperating mysteriousness of human beings in general."[9] Her view that sex should not be considered a consequential division is appealing, but not one that has yet thoroughly permeated the culture of any existing nation. Professor Catharine MacKinnon repeated the "Are women human?" question in the title of a book she published in 2007.[10] Her conclusion was "no."[11] Not because women lack humanity, but because we are deprived of it.[12]

---

people to choose an "official" gender permanently, or at all.

7. *See* Phyllis Schlafly, *A Short History of E.R.A.*, PHYLLIS SCHLAFLY REP., (Eagle Forum, St. Louis, Mo.), Sept. 1986, *available at* http://www.eagleforum.org/psr/1986/sept86/psrsep86.html.

8. Dorothy L. Sayers, *Are Women Human?: Address Given to a Women's Society, 1938*, 8 LOGOS: J. CATHOLIC THOUGHT & CULTURE 165, 177 (2005).

9. *Id.*

10. CATHARINE A. MACKINNON, ARE WOMEN HUMAN? AND OTHER INTERNATIONAL DIALOGUES (2007).

11. *Id.* at 41–42. MacKinnon explained:

> The Universal Declaration of Human Rights defines what a human being is. In 1948, it told the world what a person, as a person, is entitled to. It has been fifty years. Are women human yet?
>
> If women were human, would we be a cash crop shipped from Thailand in containers into New York's brothels? Would we be sexual and reproductive slaves? Would we be bred, worked without pay our whole lives, burned when our dowry money wasn't enough or when men tired of us,

The status of women differs from country to country, but we do not hold equal power in any of them. The attainment of true equality on a global basis as measured by economic, social, and political power is an aspirational goal that I do not expect women to achieve in my lifetime. But that does not mean that we cannot make forward progress, especially in places where women can hold jobs, own property, and vote.

When the Fourteenth Amendment was added to the Constitution in 1868 to provide for the citizenship of freed slaves,[13] the words of choice were:

> No State shall make or enforce any law which shall abridge the privileges or immunities of citizens of the United States; nor shall any State deprive any person of life, liberty, or property, without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws.[14]

Categorically subsumed within Mankind, women are etymologically included within the protected classes of "citizens" and "people."[15] Nevertheless, it took one hundred years for the United States Supreme

---

> starved as widows when our husbands died (if we survived his funeral pyre), sold for sex because we are not valued for anything else? Would we be sold into marriage to priests to atone our family's sins or improve our family's earthly prospects? Would we, when allowed to work for pay, be made to work at the most menial jobs and exploited at barely starvation level? Would our genitals be sliced out to "cleanse" us (our body parts are dirt?), to control us, to mark us and define our cultures? Would we be trafficked as things for sexual use and entertainment worldwide in whatever form current technology makes possible? Would we be kept from learning to read and write?
>
> If women were human, would we have so little voice in public deliberations and in government in the countries where we live? Would we be hidden behind veils and imprisoned in houses and stoned and shot for refusing? Would we be beaten nearly to death, and to death, by men with whom we are close? Would we be sexually molested in our families? Would we be raped in genocide to terrorize and eject and destroy our ethnic communities, and raped again in that undeclared war that goes on every day in every country in the world in what is called peacetime? If women were human, would our violation be *enjoyed* by our violators? And, if we were human, when these things happened, would virtually nothing be done about it?

Id. (footnotes omitted).

12. *See id.*

13. *See* Randy E. Barnett, *Whence Comes Section One? The Abolitionist Origins of the Fourteenth Amendment* 1 (Georgetown Pub. Law & Legal Theory Research Paper No. 10-06, 2011), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1538862.

14. U.S. CONST. amend. XIV, § 1.

15. *Id.*

Court to decide that sex discrimination could violate the Fourteenth Amendment's guarantee of equal protection.[16] Maybe women are human, at least sometimes, in some contexts, for some purposes?

Since it was incorporated into the organizing principles of the nation, the meaning of the language of the first clause of the Fourteenth Amendment has been extensively debated, and the contours of its protections have significantly evolved.[17] A Supreme Court majority announced in 1996 in *United States v. Virginia* that "neither federal nor state government acts compatibly with the equal protection principle when a law or official policy denies to women, simply because they are women, full citizenship stature—equal opportunity to aspire, achieve, participate in and contribute to society based on their individual talents and capacities."[18] Justice Antonin Scalia disagreed,[19] asserting that the Constitution takes no position on the equal protection of women.[20] More recently, and with enhanced clarity, Justice Scalia asserted in an interview with University of California Hastings College of the Law professor Calvin Massey that the U.S. Constitution does not prohibit discrimination based on sex:

> [Massey:] In 1868, when the 39th Congress was debating and ultimately proposing the 14th Amendment, I don't think anybody would have thought that equal protection applied to sex discrimination, or certainly

---

16. *See* Craig v. Boren, 429 U.S. 190 (1976) (applying intermediate scrutiny to gender-based classifications for the first time).

17. *See id.* at 190.

18. United States v. Virginia, 518 U.S. 515, 532 (1996).

19. *Id.* at 567 (Scalia, J., dissenting). Scalia wrote:

> The virtue of a democratic system with a First Amendment is that it readily enables the people, over time, to be persuaded that what they took for granted is not so, and to change their laws accordingly. That system is destroyed if the smug assurances of each age are removed from the democratic process and written into the Constitution. So to counterbalance the Court's criticism of our ancestors, let me say a word in their praise: They left us free to change. The same cannot be said of this most illiberal Court, which has embarked on a course of inscribing one after another of the current preferences of the society (and in some cases only the counter-majoritarian preferences of the society's law-trained elite) into our Basic Law. Today it enshrines the notion that no substantial educational value is to be served by an all-men's military academy—so that the decision by the people of Virginia to maintain such an institution denies equal protection to women who cannot attend that institution but can attend others. Since it is entirely clear that the Constitution of the United States—the old one—takes no sides in this educational debate, I dissent.

Id.

20. *See id.*

> not to sexual orientation. So does that mean that we've gone off in error by applying the 14th Amendment to both?
> [Justice Scalia:] Yes, yes. Sorry, to tell you that. . . . But, you know, if indeed the current society has come to different views, that's fine. You do not need the Constitution to reflect the wishes of the current society. Certainly the Constitution does not require discrimination on the basis of sex. The only issue is whether it prohibits it. It doesn't. Nobody ever thought that that's what it meant. Nobody ever voted for that. If the current society wants to outlaw discrimination by sex, hey we have things called legislatures, and they enact things called laws. You don't need a constitution to keep things up-to-date. All you need is a legislature and a ballot box. You don't like the death penalty anymore, that's fine. You want a right to abortion? There's nothing in the Constitution about that. But that doesn't mean you cannot prohibit it. Persuade your fellow citizens it's a good idea and pass a law. That's what democracy is all about. It's not about nine superannuated judges who have been there too long, imposing these demands on society.[21]

Current interpretations of the Fourteenth Amendment are unlikely to remain static in the future, as alterations are continuously proposed. For example, Senator Lindsey Graham, who represents my home state of South Carolina in the Senate and is a graduate of my employing law school, argues "that the 14th Amendment no longer serves the purpose it was designed to address and that Congress should reexamine granting citizenship to any child born in the United States."[22] If something as

---

21. Interview by Calvin Massey, Law Professor, University of California Hastings College of the Law, with Antonin Scalia, United States Supreme Court Justice (Sept. 2010), *in Legally Speaking: The Originalist*, CAL. LAW., Jan. 2011, *available at* http://www.callawyer.com/story.cfm?eid=913358&evid=1.

22. Andy Barr, *Graham: 14th Amendment Outdated*, POLITICO (Aug. 4, 2010, 10:53 AM), http://www.politico.com/ news/stories/0810/40635.html. The following conversation lays out Senator Graham's views:

> VAN SUSTEREN: All right, you're getting a lot of controversy, at least you're generating in some corners about the fact that you want to amend the 14th Amendment so that just merely being born in the United States doesn't necessarily make you a citizen. Why are you doing this?
> GRAHAM: Well, to me, I'm looking at the laws that exist and see if it makes sense today. The 14th Amendment was passed after the Civil War. Citizenship was awarded before the Civil War based on states giving citizenship. Well, after the Civil War, they were afraid that Southern states may not award citizenship to freed slaves, so they put it in the 14th Amendment that if you're naturally-born American, then you're automatically entitled to citizenship as a constitutional requirement.
>
> That made sense to me then. But now, birthright citizenship doesn't make so much sense when you understand the world as it is. You have found and I've provided you information about groups that are marketing to Chinese, and Mideastern and European families a 90-day visa package, where

fundamental as the precept that "All persons born or naturalized in the United States, and subject to the jurisdiction thereof, are citizens of the United States and of the State wherein they reside"[23] is contestable, the possibility that Scalia's view of women as neither citizens nor people could gain traction must be taken seriously. Women deserve a permanent and unambiguous instantiation of a commitment to our fundamental equality. Passage of an Equal Rights Amendment would remedy the Constitution's current failure to articulate a prohibition on sex-based discrimination so explicitly that even Justice Scalia would have to notice it is there.

We need certainty about our constitutional humanity. Though women comprise a majority of the population of the United States,[24] we do not have social, political, or economic equality with men. A general, overall preclusion of the denial or abridgement of equal rights on account of sex would be more efficient than the current piecemeal legislative approaches to eliminate the obstacles that prevent women from enjoying the same benefits of citizenship that men do.

Consider Title IX.[25] Section 1681(a) of Title IX states in pertinent part: "No person in the United States shall, on the basis of sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any education program or activity receiving Federal financial assistance . . . ."[26] Title IX was a bold, reasonably comprehensive and

---

> you come to America as a tourist. You come to a resort. You have your child at a hospital within the resort. That child is an American citizen. You turn around and leave.
>
> That to me cheapens American citizenship. That's not the way I would like it to be awarded. And you've got the other problem, where thousands of people are coming across the Arizona/Texas border for the express purpose of having a child in an American hospital so that child will become an American citizen, and they broke the law to get there.
>
> So I want to put on the table fixing immigration so we don't have a third wave in the future. We went from 3 to 12 million in the last 20 years. Twenty years from now, I don't want to have 20 million. So I think we ought to have a logical discussion. Is this the way to award American citizenship, sell it to somebody who's rich, reward somebody who breaks the law? I think we need to look at it really closely.

Interview by Greta van Susteren with Lindsey Graham, U.S. Senator, In Wash. D.C. (Aug. 3, 2010), *available at* http://www.foxnews.com/on-air/on-the-record/transcript/sen-graham-039i039m-trying-reward-american-citizenship-i039m-not-penalizing-children039.

23. U.S. CONST. amend. XIV, § 1.

24. *See* Denise I. Smith & Renee E. Spraggins, *Gender in the United States*, NATIONALATLAS.GOV, http://www.nationalatlas.gov/articles/people/a_gender.html (last visited Apr. 6, 2011).

25. Title IX, Education Amendments of 1972, 20 U.S.C. §§ 1681–1688 (2006), *available at* http://www.dol.gov/oasam/regs/statutes/titleix.htm (last visited May 5, 2011).

26. *Id.* § 1681(a).

impressively successful effort to improve women's access to educational opportunities that has been in place for nearly forty years.[27] However, it has not brought about true equality even in the context of education. Most school sports are fairly strictly segregated by sex from the time the participants are teenagers. Thanks to Title IX, girls have many (though still numerically fewer) of the same opportunities to participate in athletics as do boys,[28] so arguably both genders reap the same benefits: exercise, competition, learning the values of teamwork, tenacity, leadership, and the possibility of athletic educational scholarships. But athletic departments often get around Title IX's requirements through deceptive practices that overstate women's participation in college sports.[29]

And girls are slighted in other ways. Many girls' high school and women's collegiate teams are coached by men,[30] but rare indeed are females found coaching boys' or men's teams.[31] Female athletes coached by men are further socialized to take orders from men and reminded that coaching jobs may not be accessible to them in the future.[32] The perception that only men can be leaders or teammates is also inculcated into males whose sports experiences are woman free. Regardless of their coaches' genders, female athletes are certainly aware that their coaches are paid less, that their contests are less publicized, less often televised, attract far fewer spectators, and that they have very limited opportunities.[33] Some sports competitions, such as Olympic ski jumping, simply are not open to women.[34]

---

27. *See* Am. Ass'n of Univ. Women, *Title IX: Equity in School Athletics* 1 (2010), http://www.aauw.org/act/issue_advocacy/actionpages/upload/TitleIX_111-2.pdf.

28. *See id.*

29. *See generally* Katie Thomas, *College Teams, Relying on Deception, Undermine Gender Equity*, N.Y. TIMES, Apr. 25, 2011, http://www.nytimes.com/2011/04/26/sports/26titleix.html?_r=2&hp=&pagewanted=all (discussing how athletic departments pad rosters with unqualified participants, misleadingly count women as members of multiple athletic teams, and count as women men practicing with women's teams); Katie Thomas, *Gender Games: Answering Questions About Roster Management and Title IX*, THE QUAD: THE NEW YORK TIMES COLLEGE SPORTS BLOG (Apr. 26, 2011, 3:18 PM), http://thequad.blogs.nytimes.com/2011/04/26/gender-games-answering-questions-about-roster-management-and-title-ix/ (answering questions posed in response to her original article).

30. *See* Deborah L. Rhode & Christopher J. Walker, *Gender Equity in College Athletics: Women Coaches as a Case Study*, 4 STAN. J. C.R. & C.L. 1, 12 (2008); *see also* Sean Gregory, *Where Are the Women Coaches?*, TIME (Aug. 16, 2007), *available at* http://www.time.com/time/magazine/article/0,9171,1653648,00.html.

31. *See* Rhode & Walker, *supra* note 30, at 12.

32. *See* Gregory, *supra* note 30.

33. *Pay Inequity in Athletics*, WOMEN'S SPORTS FOUNDATION, http://www.womenssportsfoundation.org/Content/Articles/Issues/Equity-Issues/P/Pay-Inequity-in-Athletics.aspx (last visited Jan. 18, 2011).

34. Claire Suddath, *Why Can't Women Ski Jump?*, TIME (Feb. 11, 2010),

Title IX is vulnerable to efforts to diminish its power by all three branches of government.[35] Congressional representatives can try to reduce its reach. For example, in 1974, the unsuccessful Tower Amendment proposed to exempt revenue-producing sports from determinations of Title IX compliance.[36] Senator Tower tried again in 1977. The executive branch can also manipulate the reach of Title IX. President George W. Bush's administration weakened Title IX in a number of ways.[37] Judges can restrict

---

http://www.time.com/time/nation/article/0,8599,1963447,00.html; Ann Bartow, *15 plaintiffs lost their lawsuit against the Vancouver Olympic Games Organizing Committee when the British Columbia Supreme Court ruled that the decision to exclude their sport is out of the organizing committee's control*, FEMINIST LAW PROFESSORS BLOG (July 10, 2009), http://www.feministlawprofessors.com/?p=11922 (noting that ski jumping is a sport in which women can outperform men).

35. *Legislative Update Special Report: Bush Commission Weakens Title IX in Sports*, NATIONAL ORGANIZATION FOR WOMEN (Feb. 2003), http://www.now.org/issues/legislat/200302.html [hereinafter *Legislative Update*].

36. *See The Living Law*, TITLE IX, http://www.titleix.info/History/The-living-law.aspx (last visited Feb. 4, 2011).

37. *See Legislative Update*, *supra* note 35. The Secretary of Education's Opportunity in Athletics Commission conducted a study of Title IX and submitted recommended changes. *Id.* The following is critique of some of those recommendations by the National Women's Law Center:

> [(1)] While women are now 56% of undergraduates (and in some schools, women are a much larger majority, as much as 70%) one of the Commission's proposals would assume that women are 50% of the student body at all schools—regardless of the facts.
> [(2)] Another proposal would not count non-traditional students, who are overwhelmingly women; thus distorting the actual participation rates of men and women.
> [(3)] A third proposal would allow schools to pretend that they are giving female students athletic opportunities by counting "ghost slots" on teams—slots never actually filled by any female student. Still another would allow schools to pretend that they are not giving athletic opportunities to men by not counting "walk-ons" (not specifically recruited)—who are actually receiving the benefits of sports participation at the school.
> [(4)] The commission would also authorize the establishment of "variances" to permit schools to offer even fewer athletics opportunities to women under current law or new formulas.
> [(5)] The commission would allow the use of "interest surveys" to limit women's opportunities by forcing them to prove that they are interested in sports before giving them a chance to play.
> [(6)] The commission would authorize private slush funds that increase the financial support for men's teams at the expense of women's teams.
> [(7)] The commission gave a blank check to the Secretary of Education to identify "additional ways of demonstrating compliance with Title IX" that could include new ways to weaken Title IX that were not even presented to the commissioners.

the impact of Title IX by interpreting its mandates very narrowly, or by declaring it unconstitutional altogether.[38]

The educational purview of Title IX provides just one example. Women are still treated as inferior to men by the U.S. military. Women soldiers are less enthusiastically recruited and restricted from higher paying combat positions. What's more, our lesser value is communicated to all females at the cusp of adulthood when, unlike their male counterparts, they are not required to register for the draft.[39] Even opportunities for doing legal rather than martial justice are unjustly denied to women. Though we earn law degrees almost in parity with men and have done so for almost three decades,[40] women are outnumbered in the federal judiciary at every level.[41] A commitment to equality across gender identification, gonads, chromosomes, or any other maker of sex that is specifically articulated in a Twenty-Eighth Amendment to the United States Constitution would productively cut off debates about the Fourteenth Amendment and ignite engagement in projects pitched at increasing substantive equality for all persons.

Id. at 3; *see also Bush Administration Weakens Title IX: League of Fans Calls for Action to Protect Anti-Discrimination Law*, League of Fans (March 25, 2005), http://www.leagueoffans.org/titleixweakened.html.

38. *See, e.g.*, Jennifer R. Capasso, *Structure Versus Effect: Revealing the Unconstitutional Operation of Title IX's Athletics Provisions*, 46 B.C. L. REV. 825, 836 (2005).

39. Office of Public and Congressional Affairs, Selective Service System, *Backgrounder: Women and the Draft in America*, SELECTIVE SERVICE SYSTEM (July 1998), http://www.sss.gov/wmbkgr.htm.

40. *But see* Ann Farmer, *Are Young Women Turning Their Backs on Law School?*, 18 PERSP. 4, 4 (2010).

41. *See Women in Federal and State-level Clerkships*, Ctr. For Women in Gov't & Civil Soc'y, Rockefeller College of Public Affairs & Policy, University at Albany, State University of New York (Spring 2010), http://www.albany.edu/womeningov/judgeships_report_final_web.pdf.

WHY WE NEED TERM LIMITS FOR CONGRESS: FOUR IN THE SENATE, TEN IN THE HOUSE

RICHARD A. EPSTEIN*

I. INTRODUCTION: TERM LIMITS ALL OVER AGAIN

In *U.S. Term Limits, Inc. v. Thornton*,[1] a closely divided United States Supreme Court held that individual states could not impose term limits on their members of Congress.[2] On its facts, the decision was quite close, but on balance Justice Stevens had the better argument when he held that the Constitution sets the sole qualifications for election to the Senate or the House.[3] On matters that affect the entire nation, establishing uniform state requirements makes commendable sense.[4] To be sure, *U.S. Term Limits* addressed a very close and difficult question of constitutional interpretation. Many have sufficiently addressed the pros and cons of that decision, so I will not recanvass it.[5] Rather, I will argue in this short Article that the case for imposing term limits is sufficiently compelling that it should be introduced by constitutional amendment at the national level, so that it is equally binding on all states.

In adopting term limits, I think it is unwise to insist upon adopting two terms in the Senate and three terms in the House, as Arkansas did in *U.S. Term Limits*.[6] Rather, my position allows for far longer terms than the earlier proposal: four in the Senate and ten in the House. But even in this more restrained form, it should ultimately have major consequences. In order to make out this case, I shall proceed as follows. In Part II, I shall explain why both sides of the term-limits debate are mistaken in appealing to some notion of "the people." The real issue is the dangerous prisoner's dilemma game that arises whenever representatives of a national party are chosen along territorial lines.[7] In Part III, I explain why longer terms for both House and Senate members will thread the needle between excessive

---

* Laurence A. Tisch Professor of Law, New York University School of Law, Peter and Kirsten Bedford Senior Fellow, The Hoover Institution, and senior lecturer at the University of Chicago.

1. 514 U.S. 779 (1995).

2. *Id.* at 783.

3. *Id.* at 782; *see also* U.S. CONST. art. I, § 2, cl. 2; U.S. CONST. art. I, § 3, cl. 3 (enumerating the eligibility requirements to serve as a member of Congress).

4. *See* U.S. Term Limits, Inc. v. Hill, 872 S.W.2d 349, 356 (Ark. 1994).

5. *See U.S. Term Limits*, 514 U.S. at 837.

6. *See id.* at 783.

7. *See* ANATOL RAPOPORT & ALBERT M. CHAMMAH, PRISONER'S DILEMMA: A STUDY IN CONFLICT AND COOPERATION 13 (1965) ("In the game called Prisoner's Dilemma, the rational choice of strategy by both players leads to an outcome which is worse for both than if they had chosen their strategies 'irrationally.'").

turnover in government on one side and entrenched, corrupt leadership on the other.

II. THE PEOPLE AND THEIR PRISONER'S DILEMMA GAME

One of the most notable features of the opinions in *U.S. Term Limits* is the broader arguments to which each side appealed in order to demonstrate that its view was the most consistent with the highest values of democracy. As so often happens in constitutional adjudication, each side sought to wrap itself in the mantle of "the people," the highest authority in democratic politics.[8] For these purposes, it is worth mentioning the jarring conflict inside a document whose Preamble begins with the words "We the People," as if the Constitution writ large celebrated a system of popular democracy.[9] Yet the moment one turns away from the soaring rhetoric of the Preamble, the Constitution introduces a dense network of textual provisions. Many of these provisions aim to restrain popular democratic institutions, in part with a plethora of electoral obstacles that prohibit simple political majorities from impressing their will upon the public at large.[10] The two Qualifications Clauses are part of that strategy.[11]

Notwithstanding the built-in constitutional safeguards, Justice Stevens assures us that it is a "fundamental principle of our representative democracy . . . that *the people* should choose whom they please to govern them."[12] One corollary to that principle is "that the opportunity to be elected [is] open to all."[13] And later: "[W]e recognized the critical postulate that sovereignty is vested in the people, and that sovereignty confers on *the people* the right to choose freely their representatives to the National Government."[14] Justice Stevens's argument clearly tells us that, if everyone should be eligible for office, then no one can be barred by term limits. Not to be outdone, Justice Thomas referred to a different group of "the people" in his dissent:

---

8. *See infra* notes 12–15 and accompanying text.

9. U.S. CONST. pmbl.; *see, e.g.*, STEPHEN BREYER, ACTIVE LIBERTY: INTERPRETING OUR DEMOCRATIC CONSTITUTION 5 (2005) (arguing that "courts should take greater account of the Constitution's democratic nature when they interpret constitutional and statutory texts"). For criticism, see Michael W. McConnell, *Active Liberty: A Progressive Alternative to Textualism and Originalism?*, 119 HARV. L. REV. 2387, 2394 (2006) (reviewing STEPHEN BREYER, ACTIVE LIBERTY: INTERPRETING OUR DEMOCRATIC CONSTITUTION (2005)).

10. *See* BREYER, *supra* note 9, at 29 ("The Framers' goal was to 'secure the public good and private rights against the danger of [factionalism], and at the same time to preserve the spirit and form of popular government.'").

11. *See* U.S. CONST. art. I, § 2, cl. 2; U.S. CONST. art. I, § 3, cl. 3.

12. U.S. Term Limits, Inc. v. Thorton, 514 U.S. 779, 793 (1995) (quoting Powell v. McCormick, 395 U.S. 486, 547 (1969)) (emphasis added) (internal quotation marks omitted).

13. *Id.* at 793–94 (emphasis added).

14. *Id.* at 794.

> I see nothing in the Constitution that precludes the people of each State (if they so desire) from authorizing their elected state legislators to prescribe qualifications on their behalf. If *the people* of a State decide that they do not trust their state legislature with this power, they are free to amend their state constitution to withdraw it. This arrangement seems perfectly consistent with the Framers' scheme.[15]

In my view, both sides of this debate missed the essential element by treating an appeal to "the people" as the ultimate source of legitimacy. The real difficulty associated with term limits arises in quite a different fashion. In its crudest form, the objection is simply a prisoner's dilemma game that fosters an odd sense of inversion in our constitutional structure.[16] For example, in the presidential election, it is the people at large who must choose only one person to serve as President of the entire nation. At this point, much can be said for the view that there is no reason why the people cannot elect the President for as many terms as they choose. Nonetheless, in the aftermath of Franklin D. Roosevelt's nearly four-term presidency, the Twenty-Second Amendment introduced term limits for the President.[17] Today there are no term limits for the Senate or the House. There is only a serious prisoner's dilemma game because "the people" of one state cannot vote out of office the longstanding senators or representatives of *other* states and *other* districts.

All elected state officials face some conflict in balancing their state interests with their national responsibilities. That the people of one state cannot upend another state's senators or representatives is troublesome because some mechanism is necessary to mediate that conflict.[18] This difficulty is one that the current system of democratic politics cannot address, let alone resolve. As a matter of brute public choice theory, reelection is one key constraint that weighs heavily on just about all members of Congress. To gain reelection, members of Congress must meet a powerful territorial constraint: garnering support from a majority of voters in their district.[19] As for campaign support, outside sentiments matter only indirectly.

The territorial nature of our political system directs elected officials to look locally even though their public duties extend nationally.[20] At this

---

15. *Id.* at 883 (Thomas, J., dissenting) (emphasis added).

16. *See* BREYER, *supra* note 9, at 28 ("[T]he Constitution's structural complexity . . . [seeks] to produce a form of democracy that would prevent any single group of individuals from exercising too much power . . . .").

17. *See* U.S. CONST. amend. XXII, § 1.

18. *See* BREYER, *supra* note 9, at 29.

19. *See* Gary S. Becker, *A Theory of Competition Among Pressure Groups for Political Influence*, 98 Q.J. ECON. 371, 371 (1983) ("The economic approach to political behavior assumes that actual political choices are determined by the efforts of individuals and groups to further their own interests.").

20. *See* BREYER, *supra* note 9, at 28. In the post-revolutionary United States, "'the

point, few politicians can resist such a great temptation. They hope to obtain key committee positions, enter into strategic alliances, and gain enough local publicity to persuade their constituents to support their reelection campaigns. To accomplish these goals, elected officials are tempted to funnel benefits back home that are disproportionate to their district's population.[21] In this context, airports are for construction jobs—not national transportation—and paving "roads to nowhere" leads to political success. This strategy is feasible because of the virtually nonexistent constitutional constraints on the ability to redistribute wealth through various permutations of the taxing and spending power.[22]

Length in office offers a huge advantage to those seeking localized benefits. Although seniority does not decide everything in Congress, the clout that accompanies long service in office remains a prominent ingredient in obtaining political power. Constituents understand this point well. In their role as "the people," they are quite happy to reelect those officials who bring home the bacon, even if it comes at the expense of the nation's overall welfare.[23] There is no magic formula in appropriations that can balance this tendency. Whether in the form of a new military base and processing facility or a new regional office for an administrative agency, even so-called public goods must have a designated location. The longer one remains in office, the more constituent goodies the diligent politician can secure, thereby reducing the voters' incentive to vote that official out of office. This prisoner's dilemma game for the House and Senate is a true scourge on national politics.

As I have lamented on more than one occasion, there was no way that I could vote against Robert Byrd or Jesse Helms while a resident of Illinois, even though removing them from public office would benefit me far more than choosing my own senator.[24] As these individuals gained clout, they could certainly direct more goodies to their home states, while the less powerful Illinois senators could not match their entrepreneurial activity. As a local citizen, I envied senators and representatives from other regions who had the clout that my elected officials so lacked. As an academic theorist,

---

great objects' of society were 'sacrificed constantly to local views.'" *Id.* Groups with "divergent social, economic, and religious interests" tended to "choose representatives . . . for their willingness to act solely to advance [their] particular interests." *Id.*

21. *See* Burton A. Abrams & William R. Dougan, *The Effects of Constitutional Restraints on Governmental Spending*, 49 PUB. CHOICE 101, 102 (1986) ("In equilibrium, successful politicians choose the set of expenditures . . . and taxes . . . that maximize political support.").

22. *See* U.S. CONST. art. I, § 8, cl. 1.

23. *See* Abrams & Dougan, *supra* note 21, at 102 ("[R]elatively influential groups will be net beneficiaries of government spending, while the members of relatively weak groups will tend to pay more in taxes than they receive in benefits from spending.").

24. JOINT COMM. ON PRINTING, CONGRESS OF THE UNITED STATES, BIOGRAPHICAL DIRECTORY OF THE UNITED STATES CONGRESS 1774–2005, H.R. DOC. NO. 108-222, at 762, 1232–33 (2d Sess. 2005).

however, the more pressing point was not to make sure that my district joined the list of preferred locales, but rather to make sure that fewer districts carried that coveted title in the first place.

Short of term limits, I do not think that there is any conceivable way in which this could be done, at least within any sensible time horizon. In my own distinctive take on substantive constitutional law, any government program that by taxation and expenditure, or by direct regulation, worked a net transfer from A to B should count as a taking that must be enjoined if not compensated.[25] If this regime were put into place, the opportunities for territorial manipulation would surely be reduced as big-ticket items like ethanol subsidies for the good state of Iowa come off the table.[26] Nonetheless, there is always unevenness in the distribution of traditional public goods, which have to be located somewhere. Perhaps good service rules would mute the scope of competition for these plums. For instance, in some cases, as with the closing of military bases, pressures get so strong that an independent panel must take care of the task to ensure success.[27] However, the long and short of it is that reforms of this sort could easily take decades to be introduced. Even the agricultural subsidies from the New Deal are still very much with us as the key provisions of the Agricultural Adjustment Acts continue in force to this very day.[28] What is needed is the short-term clout that term limits can impose.

In similar fashion, reform of the budget process can do little to stop the skew in the distribution of public goods.[29] Even a balanced budget

---

25. *See* RICHARD A. EPSTEIN, TAKINGS: PRIVATE PROPERTY AND THE POWER OF EMINENT DOMAIN 286 (1985).

26. *See* Roberta F. Mann & Mona L. Hymel, *Moonshine to Motorfuel: Tax Incentives for Fuel Ethanol*, 19 DUKE ENVTL. L. & POL'Y F. 43, 72–73 (2008) (discussing how presidential candidates have used the importance of ethanol to Iowa's economy as a political tool).

27. *See* Base Closure and Realignment Act, 10 U.S.C. § 2687 (2006) (creating an independent commission to cut government costs by closing military facilities); Natalie Hanlon, *Military Base Closings: A Study of Government by Commission*, 62 U. COLO. L. REV. 331, 333–40 (1991) (following the development of the Commission for the Base Closure and Realignment Act). On the one hand, members of Congress fought to get and keep military bases in their own jurisdictions to funnel defense money into the state economy. *Id.* at 333–34. However, in 1988 Congress faced budget pressure and authorized the independent commission to make reports on base closings. *Id.* at 336. Members of Congress were torn between fighting for military money in their respective states and cutting the budget costs to appease voters. *Id.*

28. *See* Guadalupe T. Luna, *The New Deal and Food Insecurity in the "Midst of Plenty"*, 9 DRAKE J. AGRIC. L. 213, 217, 240 (2004) (noting both the legacy of the New Deal and its particular impact on current agricultural subsidies).

29. *See* Elizabeth Garrett, *Rethinking the Structures of Decisionmaking in the Federal Budget Process*, 35 HARV. J. ON LEGIS. 387, 387 (1998) ("The federal government's budget decisions inevitably involve trading the demands of some groups against those of others."). *See generally* Tim Westmoreland, *Standard Errors: How Budget Rules Distort Lawmaking*,

amendment, which attempts to limit aggregate expenditures, cannot prevent goods' skewed distribution across different constituencies.[30] Further, the power of the small states in the Senate, which remains a constitutional constant, gives them a long-term advantage in this game. Public criticism may work to slow down the skew for a short while, but the heavy hitters can lay low while the public storm rages and reassert their traditional prerogatives after the storm passes. Therefore, none of these options will work to constrain factionalism. Term limits offer the best prospect of reform.

III. LONGER TERM LIMITS

The use of term limits should have a desirable short-term effect. Rotation in and out of office, an ancient practice, was designed to limit the corrosive effects of time in public service.[31] A constitutional amendment for term limits could push this concept along mightily by refusing to exempt entirely those individuals who now hold public office. Under this new term-limit amendment, incumbent senators who are over the four-term limit would be able to finish their current terms and hold office for one additional term. A similar solution would hold in the House. Those who have been in office under five terms in the House would be immediately subject to the rule. However, those House members who have served between five and ten terms would get one additional term. Consequently, the outward procession of Congress would start sooner rather than later.

The point of longer terms than those adopted by Arkansas in *U.S. Term Limits* is clear enough. Three terms in the House, is, for sure, too short.[32] In effect, a three-term House limit would require a rotation in and out of office of one-third of the House membership every two years, leading to massive disruption in government operations and the loss of any acquired expertise of particular House members.[33] Undermining the institutional memory of

---

95 GEO. L.J. 1555 (2007) (explaining how the federal budget process has created skewing of congressional choices).

30. *See* Theodore P. Seto, *Drafting a Federal Balanced Budget Amendment That Does What It Is Supposed to Do (And No More)*, 106 YALE L.J. 1449, 1461–62 (1997) (discussing how the scarcity of public goods supports one rationale behind a balanced budget).

31. *See* Mark P. Petracca, *Do Term Limits Rob Voters of Democratic Rights? An Evaluation and Response*, 20 W. ST. U. L. REV. 547, 564 (1993) ("Throughout history, from the Athenian and Roman experiments with democracy and the writings of the English Commonwealthmen in the 17th and 18th centuries to advocacy by America's revolutionaries, the principle of rotation in office was an institutional feature of a legislative body.") (footnotes omitted); *see also id.* at 563 (linking the historical concept of rotation in office to the modern concept of term limits).

32. *See* 141 CONG. REC. 9723 (1995) (proposed amendment for three-term limit in the House).

33. *See* U.S. CONST. art. I, § 2, cl. 1 (requiring members of the House to be chosen every second year).

the House could easily destabilize short-term politics and result in inordinate influence by key staff members, who could take over the reins of power. When the House term limit moves up to ten terms from three, the rate of rotation in equilibrium now slows down. At each election cycle, ten percent of members have to leave, and probably another five percent or so will leave for a variety of reasons, including defeat at the polls. Yet the members of the House that remain should be able to carry on in an orderly fashion. A similar set of arguments applies to the Senate, where the rate of turnover is only slightly higher.[34]

To get some idea of what congressional term limits could do, just think of what they would have done to the careers of some of the most notable figures in both houses of Congress. Start with the barons of the House who would have lost office under these rules. That list includes just about all of the Speakers of the House:[35] Sam Rayburn, who served forty-eight years;[36] John McCormack, who served for forty-three years;[37] Jim Wright, who clocked in at thirty-four years;[38] and Nancy Pelosi, who is still active with twenty-three years in office and assumed the speakership in her twentieth year in office.[39] A quick look on the Senate side shows such notables as Lyndon B. Johnson with twelve years each in the House and in the Senate, where he served as majority leader for his last six years.[40] His rapid rise in the Senate was due, in part, to his earlier experience in the House, and I see no reason to be unduly worried about the switch between the two chambers. Few can make it from the House to Senate, and few will choose to make the return trip from Senate to House. The two longest-serving senators were Robert Byrd[41] and Strom Thurmond,[42] each with close to sixty years in office. Another notable senator, Ted Kennedy, served for forty-seven years.[43] Ten terms could make a huge difference in the House. The four-term limit in the Senate could also transform that body, in general, for the better.

Note that under the current rules, these individuals did not come to power in either the House or Senate during the early part of their careers, in

---

34. *See* Aaron-Andrew P. Bruhl, *Burying the "Continuing Body" Theory of the Senate*, 95 IOWA L. REV. 1401, 1436–37 (2010) (comparing rate of reelection for incumbents in the House and the Senate).

35. *See generally House History: Speakers of the House*, OFF. CLERK U.S. HOUSE REPRESENTATIVES, http://clerk.house.gov/art_history/house_history/speakers.html (last visited Jan. 26, 2011) (listing House Speakers from 1789 to 2011).

36. JOINT COMM. ON PRINTING, BIOGRAPHICAL DIRECTORY OF THE UNITED STATES CONGRESS: 1774–2004, H.R. DOC. NO. 108-222, at 1787–88 (2005).

37. *Id.* at 1533.

38. *Id.* at 2203.

39. *Id.* at 1718.

40. *Id.* at 1339.

41. *Id.* at 762.

42. *Id.* at 2045.

43. *Id.* at 1371–72.

part because senior officials stood in their way. Yet that is exactly why there is much to be said for having a rapid rotation so that abler people can rise more quickly to the top. In addition, the mindset of those people who know that they are about to leave, like those who are just entering, is likely to be somewhat less statist than might otherwise be the case. A system of rotation ensures, moreover, that the government will not become a geriatric society in which a disproportionate influence rests in the hands of individuals who are likely to plump hard to Social Security and Medicare entitlements. The new brush will sweep clean.

The question then arises whether imposing term limits has the downside of driving out individuals from government who belong there on their individual merits. No one could doubt that this outcome would occur in some cases. But it is a mistake to think that the issue is one that should dominate the discussion on this matter. Like so many other questions, this one boils down to a clear assessment of two kinds of error. The first of these is the error of keeping people in office who should be out. In my view, the longer the passage of time, the more likely it is that this form of error will dominate. Worse still, the harm that comes from continued excessive influence based on seniority only gets worse with age. But on the other side, those individuals who are forced out of office in their prime still have useful lives ahead of them. They can run for other political offices or undertake other useful endeavors. In short, there are ways to correct through individual action the mistakes that arise when people are forced out of office, but it is far harder to take corrective steps against entrenched government representatives with the capacity to barricade themselves in office.

CONCLUSION

One notable feature about the term-limits movement is that virtually all of its ardent supporters hold a small government, libertarian orientation. I can recall attending meetings of the Cato Institute,[44] where a report on term limits was a standard agenda item, about which there was no substantive disagreement. The compliment is returned on the other side of the aisle. Most of the term limit opponents are far more comfortable with the large welfare state. I doubt that this is a coincidence. Indeed, I think that the support for the term limit movement is part and parcel of a small-government approach that should be defended, especially in these hard times, on the grounds that, given where we now stand, more government is worse government.

The great tactical mistake of the earlier generation of term-limit supporters is that they often had a not-so-covert desire to cripple government. I do not. Even limited governments have huge amounts of

44. *See* CATO INSTITUTE, http://www.cato.org (last visited May 1, 2011).

constructive things that they can and must do. The case for term limits is much like the case for exit rights under federalism.[45] It does not necessarily lead to any particular substantive result, but it creates an environment that, on balance, will tend to shrink government when that change is most needed. That gamble is, of course, no different from the original gamble of the Framers that devices intended to reduce government power would do better, on average, than those devices intended to expand it. The air is now out of the term-limits movement,[46] but as the reaction to the current malaise increases, voters from all parties should come to realize that a new broom that sweeps relentlessly clean is yet another of the structural protections that form the nondemocratic backbone of a sound and stable democratic society, which cannot survive under a simple regime of majority rule.

45. *See generally* Richard A. Epstein, *Exit Rights Under Federalism*, 55 LAW & CONTEMP. PROBS. 147 (1992).

46. *See* Elizabeth Garrett, *Term Limitations and the Myth of the Citizen-Legislator*, 81 CORNELL L. REV. 623, 624 (1996) (calling the Supreme Court's decision in *U.S. Term Limits, Inc. v. Thornton* "a substantial roadblock" to the term-limits movement).

THE EXECUTION SHOULD BE TELEVISED: AN AMENDMENT MAKING EXECUTIONS PUBLIC

DAVID LAT* & ZACHARY SHEMTOB**

"What this country needs is for public executions to be reinstated."
Justice Lewis F. Powell, Jr., The Brethren (1979)

At the time of the American Revolution, executions were conducted in public.[1] Our proposed constitutional amendment would make them public once again. Making executions public might seem intuitively troubling or even morally archaic. Reinstating public executions, however, would offer a profound affirmation of democratic transparency and accountability.

In a genuine democracy, matters of such importance as the death penalty should be clearly and firmly supported or repudiated. But because the ultimate punishment occurs behind closed doors,[2] it is all too easy for citizens to overlook the issue entirely. Returning executions to the public sphere would force Americans either to openly endorse or firmly reject a sanction that many find all too easy to ignore.

CAPITAL PUNISHMENT'S IMPORTANCE

Capital punishment remains a matter of great concern in the nation's criminal justice system. Over 1,100 persons have been executed since the death penalty's reinstatement in 1976 in *Gregg v. Georgia*,[3] and as of January 2010, 3,261 inmates sat on death row in thirty-six states (excluding the U.S. government and military).[4] Capital punishment has also cost billions of dollars and will likely continue to do so. California has estimated that death row inmates cost taxpayers $114 million more per year than those serving a life sentence,[5] and Texas has reported death penalty cases

---

* David Lat is the founder and managing editor of the legal website Above the Law.

** Zachary Shemtob is an Assistant Professor in Criminal Justice and Criminology at Central Connecticut State University.

1. *See* Nicholas Levi, *Veil of Secrecy: Public Executions, Limitations on Reporting Capital Punishment, and the Content-Based Nature of Private Execution Laws*, 55 FED. COMM. L.J. 131, 134 (2002).

2. *See infra* note 18 and accompanying text.

3. 428 U.S. 153, 207 (1976).

4. *See* CRIMINAL JUSTICE PROJECT, NAACP LEGAL DEF. AND EDUCATIONAL FUND, DEATH ROW U.S.A. WINTER 2010, 1, 35–36 (2010) *available at* naacpldf.org/files/publications/DRUSA_Winter_2010.pdf.

5. *See The High Cost of the Death Penalty*, DEATH PENALTY FOCUS, http://www.deathpenalty.org/article.php?id=42 (last visited May 5, 2011) (citing Rone Tempest, *Death Row Often Means a Long Life; California Condemns Many Murderers, but Few Are Ever Executed*, L.A. TIMES, Mar. 6, 2005, at B1).

generate an average cost of $2.3 million, or about three times the amount of imprisoning someone for life.[6] While such costs are always considerable, they are especially significant in a time of economic difficulty, when state budgets are more strained than ever.[7]

For opponents, cost is not the only issue. The death penalty consumes an incredible amount of judicial resources because capital cases are often subjected to a lengthy and Byzantine appeals process.[8] Capital punishment also implicates fundamental concerns of fairness, along both socioeconomic and racial lines. Poor defendants are more likely to be executed than wealthy ones, and a defendant is most likely to receive the death penalty if he is African-American and the victim is Caucasian.[9] Executions have also exposed the United States to international criticism; they are banned by the United Nations Charter and prohibited within the European Union.[10]

At the same time, capital punishment is deeply important to its defenders. To proponents of the death penalty, it powerfully embodies "society's recognition of the sanctity of human life."[11] It is, to many of its supporters, the only appropriate punishment for the most heinous of crimes: "As Lord Justice Denning argued in 1950, 'some crimes are so outrageous that society insists on adequate punishment, because the wrong-doer deserves it, irrespective of whether it is a deterrent or not.'"[12]

---

6. *See* C. Hoppe, *Executions Cost Texas Millions*, DALLAS MORNING NEWS, Mar. 8, 1992, at A1; *see also* AM. CIVIL LIBERTIES UNION, THE HIGH COSTS OF THE DEATH PENALTY 4–8, *available at* www.nacdl.org (click "search," then enter article title) (last visited May 5, 2011).

7. *See* Michael Cooper and Mary Williams Walsh, *Mounting State Debts Stoke Fear of a Looming Crisis*, N.Y. TIMES, Dec. 5, 2010, at A1. Total estimated state budget shortfalls are projected to reach $134 billion in 2012. *See* Iris Lac & James Horney, *House GOP Leaders' Plan Would Slash Funds for State and Local Services, Slow Economic Recovery*, CENTER ON BUDGET AND POLICY PRIORITIES (November 11, 2010), http://www.cbpp.org/cms/index.cfm?fa=view&id=3323; *see generally* NICHOLAS JOHNSON ET AL., CTR. ON BUDGET AND POLICY PRIORITIES, AN UPDATE ON STATE BUDGET CUTS (2010), *available at* www.cbpp.org/3-13-08sfp.pdf (examining how the recession has affected state budgets).

8. *See* AM. CIVIL LIBERTIES UNION, *supra* note 6, at 3–10 (exploring the financial burden that the death penalty imposes on state and local governments and arguing that life imprisonment is much less costly and burdensome).

9. *See* EVAN J. MANDERY, CAPITAL PUNISHMENT: A BALANCED EXAMINATION 379–80 (2004); *see also* Baze v. Rees, 553 U.S. 35, 85 (2008) (Stevens, J., concurring in the judgment) (noting "the risk of discriminatory application of the death penalty," and claiming that "the Court has allowed it to continue to play an unacceptable role in capital cases").

10. *See, e.g.*, EU Policy on the Death Penalty, EUR. UNION, http://www.eurunion.org/legislat/deathpenalty/eumemorandum.htm (last visited May 5, 2011).

11. *See* Matt Chandler, *Hatch to Speak Here Monday*, BUFFALO LAW JOURNAL (March 25, 2010), *available at* http://www.lawjournalbuffalo.com/news/article/current/2010/03/25/102187/hatch-to-speak-here-monday (quoting Senator Orrin Hatch of Utah).

12. *Baze*, 553 U.S. at 80 (Stevens, J., concurring in the judgment) (quoting Gregg v. Georgia, 428 U.S. 153, 184 n.30 (1976)).

Like abortion, capital punishment is an issue on which advocates and opponents have been able to find little common ground.[13] While proponents praise the death penalty as advancing justice, abolitionists find it abhorrent, some going so far as to label it as "halfway" to the Holocaust.[14] Proponents like Wesley Lowe and opponents such as Sister Helen Prejean not only make clashing arguments, but they also seem to employ entirely different moral vocabularies.[15]

OUT OF SIGHT, OUT OF MIND

For an issue that engenders such intense conflict, capital punishment is remarkably hidden from public view. While states such as New York had bills privatizing executions as early as 1835,[16] the last public hanging in the United States occurred as recently as 1936.[17] Since then, this most serious of sanctions has been imposed almost exclusively behind closed doors; only the condemned and victims' families, along with a select group of media representatives, are allowed to attend.[18] As further testament to the death penalty's secrecy, executions almost always occur between midnight and sunrise, and the executioners' identities are held in the strictest confidence.[19]

The intense privacy of executions, combined with their relatively rarity (2009 saw a total of fifty-two executions out of over 3,000 individuals on death row),[20] effectively walls them off from public scrutiny. Despite their strong support for capital punishment,[21] Americans are also strikingly fickle when forced to give the ultimate sanction greater thought. Although a majority of Americans continue to support capital punishment, this often

---

13. *See* Levi, *supra* note 1, at 132 ("Few issues in America spark more robust debate and disagreement than capital punishment.").

14. *See* Roger Ebert, *Mr. Death: The Rise and Fall of Fred A. Leuchter, Jr.*, CHICAGO SUN TIMES, Feb. 4, 2000, *available at* http://rogerebert.suntimes.com/apps/pbcs.dll/article?AID=/20000204/REVIEWS/2040302/1023.

15. *Compare* Wesley Lowe, *Pro Death Penalty Webpage* (last updated Jan. 17, 2011), http://www.wesleylowe.com/cp.html, *with* Sister Helen Prejean, *News From Sister Helen*, (last visited May 5, 2011), http://www.prejean.org/NewsFrom.html.

16. *See* JOHN D. BESSLER, DEATH IN THE DARK: MIDNIGHT EXECUTIONS IN AMERICA 47 (1998).

17. *Id.* at 4 (noting that the last public execution took place in Owensboro, Kentucky).

18. *See* STUART BANNER, THE DEATH PENALTY: AN AMERICAN HISTORY 161–63, 168 (2003); BESSLER, *supra* note 16, at 96–97.

19. *See* BESSLER, *supra* note 16, at 81–82, 151.

20. *See Statistics*, ANTIDEATHPENALTY.ORG, www.antideathpenalty.org/statistics.html (last visited May 5, 2011).

21. *See* Frank Newport, *In U.S., Two-Thirds Continue to Support Death Penalty*, GALLUP, Oct. 13, 2009, *available at* www.deathpenaltyinfo.org/documents/GallupPoll1009.pdf (finding continued support for the death penalty: 65% of Americans support its use while only 31% of Americans oppose it).

changes when they are asked to contemplate other alternatives such as life without parole.[22] This research suggests that capital punishment, despite its great controversy and severity, remains an issue of considerable public ignorance.[23]

Such ignorance, and the social apathy it engenders, is an affront to democracy. Citizens in a democracy must have some say over the policies that govern them. While electing representatives may blunt somewhat the full possibility of this, legislators are generally expected to carry out the will of their constituents. Integral to this process is that citizens fully understand what they are voting for, namely, the policies their representatives seek to institute. In order to understand a particular policy, however, one must have some conception of its actuality. Merely pulling a lever and moving on is simply not enough; having some idealized notion of how something works is wholly different from watching it work. And herein lies the problem: Veiling capital punishment from the general citizenry has reduced it to a mere abstraction.

If citizens grant the State the right to impose the most serious sanction of which it is capable, they must fully realize what this process entails. This involves not only abstract support for the death penalty but the opportunity to see it in action. Sporadic media accounts are inadequate—there is a substantial difference between witnessing an execution and comfortably reading about it after the fact. If the majority of citizens cannot handle such a prospect, then it may be time to revoke the State's license to kill.

The death penalty is far too important a policy to be imposed by default or as a result of societal inertia. But this may be exactly what is happening in this nation, as Justice Stevens argued in his concurrence in *Baze*:

> [C]urrent decisions by state legislatures, by the Congress of the United States, and by this Court to retain the death penalty as a part of our law are the product of habit and inattention rather than an acceptable deliberative process that weighs the costs and risks of administering that penalty against its identifiable benefits . . . .[24]

---

22. *See* John K. Cochran & Mitchell B. Chamlin, *Can Information Change Public Opinion? Another Test of the Marshall Hypothesis*, 33 J. CRIM. JUST. 573, 573 (2005) (study with overall mixed findings, but reporting some evidence that suggests that gains in knowledge about the death penalty should be associated with reductions in death penalty support); Carol S. Steiker, *The Marshall Hypothesis Revisited*, 52 HOW. L.J. 525, 553 (2009) (examining the implications of the Marshall hypothesis and finding support for it in "all kinds of places—from the laboratory, to the wider world, to the Supreme Court itself"). Steiker explains how Justice Marshall's opinion in *Furman v. Georgia*, 408 U.S. 238 (1972), where he predicted that the majority of Americans would oppose the death penalty if fully informed, became known as the "Marshall hypothesis." *Id.* at 527–28.

23. *See* Cochran & Chamlin, *supra* note 22, at 573; Steiker, *supra* note 22, at 525.

24. Baze v. Rees, 553 U.S. 35, 78 (2008) (Stevens, J., concurring in the judgment).

Continued use of the death penalty may well be appropriate—but it should be sustained as the result of a deliberative process that includes an understanding of what an execution entails.

DISCONTENTS

One can imagine a number of objections to making executions public. Perhaps the most obvious is that publicizing executions would profoundly shock and disturb the American public.[25] Yet this may be the very point. If citizens in a democracy cannot handle what is being done in their name, perhaps it ought not to be done.

Other critics may attack the prospect of public executions from the opposite end, fearing that the public may actually enjoy witnessing such spectacles.[26] But if the public celebrates such events, then this is something we need to honestly acknowledge and openly contemplate. In the same vein, there is a fear of the so-called brutalization effect—the idea that public executions may actually make people more violent. Although no empirical proof exists to support such speculation,[27] fear of the potential public reaction should hardly stifle genuine democratic debate.

Critics might also contend that children could fatally copy what they see, failing to understand the full consequences of an execution. Ideally, parents would shield their children from these events, just as they would from any violent content, but this prospect is obviously somewhat unrealistic. Regardless, public executions, at least in the eyes of children, would seemingly have no more impact than violent content in general. Indeed, this practice could potentially make children more aware that violence is not a trivial event, just as witnessing an actual death radically differs from the cartoonish violence to which children are routinely exposed.

Finally, supporters of capital punishment will likely contend that public executions are a crafty attempt to end the death penalty through the back door. But while capital punishment's opponents have been considerably more open to the idea than its supporters, our intention is not one of abolition.[28] The public's acceptance or rejection of capital punishment after confronting it so starkly would fulfill our objective in equal measure.

---

25. *See generally* CHRISTOPHER S. KUDLAC, PUBLIC EXECUTIONS: THE DEATH PENALTY AND THE MEDIA 11, 17, 103–104 (2007) (describing public executions and media publicity of executions).

26. *See generally* MARTHA C. NUSSBAUM, HIDING FROM HUMANITY: DISGUST, SHAME, AND THE LAW 242–44 (2004) (finding public penalties troubling).

27. *See* David R. King, *The Brutalization Effect: Execution Publicity and the Incidence of Homicide in South Carolina*, 57 SOC. FORCES 683, 687 (1978) (finding no support for the hypothesis that publicity about executions stimulates homicide).

28. *See, e.g.*, BESSLER, *supra* note 16, at 180–96, 210–11.

A CONSTITUTIONAL MATTER?

Only a constitutional amendment can fully realize this proposal. Although public executions have yet to face Supreme Court review, "no court—state or federal—has ever held that there is a constitutional right to film and to broadcast executions."[29] While some have argued that the First Amendment protects the right to televise the imposition of the death penalty,[30] and the concerns underlying the First Amendment certainly support our proposal, this issue is different than the one at hand. We are not considering whether capital punishment *may* be made public, but asserting that it *must* be made public.

Neither individual states nor the federal government are likely to adopt such a bold position alone. Any state that did so unilaterally would surely face incredible scrutiny, especially in a time of such economic and social turmoil. Instituting such a controversial policy would also result in a host of legal challenges, making its enactment feasible only on a national scale. Equally important, simple legislation, even if it were passed, would lack the political and legal force of a formal constitutional amendment.

The form that public executions should actually take is another matter, one that lies well beyond the scope of the present paper. Just as states enjoy significant discretion in how they implement the death penalty under the Eighth Amendment,[31] here too they should be afforded discretion in terms of how to make executions "public."

Perhaps executions could be performed in large public squares, as they were in colonial times, although the logistics would undoubtedly be challenging. A more practical idea would be the use of closed-circuit television or streaming broadcasts over the internet. Regarding privacy interests, executioners' faces could be digitally covered or altered. The technological advances of the digital age make it easier than ever to make executions public while reducing logistical difficulties or the prospect of public disorder.[32]

---

29. Levi, *supra* note 1, at 144; *see also* Houchins v. KQED, Inc., 438 U.S. 1, 11 (1978) (quoting Pell v. Procunier, 417 U.S. 817, 834 (1974)); *accord* Saxbe v. Wash. Post Co., 417 U.S. 843, 850 (1974); Garrett v. Estelle, 556 F.2d 1274, 1279 (5th Cir. 1977); KQED, Inc. v. Vasquez, No. C-90-1383 RHS, 1995 WL 489485, at *1 (D. Cal. Aug. 1, 1991).

30. *See, e.g.*, Levi, *supra* note 1, at 132; Jef I. Richards & R. Bruce Easter, *Televising Executions: The High-Tech Alternative to Public Hangings*, 40 UCLA L. REV. 381, 382–83 (1992); Jerome T. Tao, Note, *First Amendment Analysis of State Regulations Prohibiting the Filming of Prisoner Executions*, 60 GEO. WASH. L. REV. 1042, 1045 (1992).

31. *See* Baze v. Rees, 553 U.S. 35, 62 (2008) (rejecting petitioners' claim that Kentucky was constitutionally required to follow a particular lethal injection protocol).

32. *See* Levi, *supra* note 1, at 139 ("By the 1820s and 1830s public executions were a violent and commercially exploited spectacle that often led to drunken riots.").

The practicalities of making executions public—to be worked out by individual states and subject to judicial review—are ultimately less important than the principles at stake. Even if people refused to watch the newly public executions, the sheer publicity of such events would force them into public consciousness and trigger tremendous debate in newspapers, on blogs, and around kitchen tables.

This debate is long overdue. As Justice Stevens stated in *Baze*, "The time for a dispassionate, impartial comparison of the enormous costs that death penalty litigation imposes on society with the benefits that it produces has surely arrived."[33] Although Justice Stevens was speaking specifically about litigation surrounding the death penalty, his words ring just as true for the death penalty itself.

DEMOCRATIZING DEATH

Some will surely express intuitive revulsion at such an unorthodox idea. The idea of streaming executions live over the internet might engender the kind of shock as having a "human sacrifice channel" on television.[34]

While such reactions are certainly understandable, it is equally important to step back and rationally consider the issue at hand. Walling off executions has transformed these events into punitive abstractions, allowing the public to blissfully ignore society's most serious sanction. A more genuine accounting may result in discomfort or derision, but this is ultimately part of living in an authentic democracy. Forcing executions into the public sphere would express ideals continually professed but rarely fulfilled, inserting a measure of reality into a society unwilling to look one of its most profound practices in the eye.

A constitutional convention represents a moment of heightened consciousness for the body politic, in which the system's actors step away from their day-to-day concerns and devote their attention to more fundamental principles. Similarly, making executions public would heighten the citizenry's consciousness of the death penalty. Either the people will reject the death penalty or they will embrace it—but at least their choice, right or wrong, would not be made unthinkingly.

---

33. *Baze*, 553 U.S. at 81 (Stevens, J., concurring in the judgment).

34. *See* Transcript of Oral Argument at 46, United States v. Stevens, 130 S. Ct. 1577 (2010) (No. 08-769), *available at* http://www.supremecourt.gov/oral_arguments/argument_transcripts.aspx (in which Justice Alito raises the possibility of human sacrifices being broadcast "[l]ive, pay per view, you know, on the human sacrifice channel.").

AFTERWORD: FULL OF SOUND AND FURY BUT SIGNIFIYING RELATIVELY LITTLE?

SANFORD LEVINSON*

I commend the editors of the *Tennessee Law Review* for commissioning this symposium. From my perspective, it is almost irrelevant whether I agree with all of the proposals (which I certainly do not). Rather, I think it is absolutely crucial that Americans develop a more critical stance toward what I regard as a dangerously flawed and dysfunctional Constitution, and one way of doing that is precisely to ask the simple question, "What kinds of constitutional amendments would make it better?" I disagree vehemently with those who suggest that our national Constitution is sufficiently perfect as it is, so that talk of amendment is necessarily "wacky," if not unpatriotic.

Whether paradoxically or not, denunciations of what Professor Kathleen Sullivan has memorably labeled "constitutional amendmentitis"[1]—as if the very proposal of constitutional amendments constituted a disease—have tended to come more from political liberals than from conservatives, even though what I have called, not altogether flatteringly, "constitutional veneration" is much more likely, these days, to be found on the political right.[2] The liberal aversion to amendment may be explained simply by the fact that it has been the political right, over the past generation, that has been far more likely to suggest constitutional amendments, whether to balance the budget—the particular subject of Sullivan's scorn—to ban flag burning, or to prohibit same sex marriage.[3] Still, it is telling that rather than simply confront these suggestions on their merits, liberals have chosen to condemn the very idea of amendment.[4] This impulse is itself terrible for the body politic; a major consequence is to reinforce the almost infantile adulation of the (unamended) Constitution that is such an unfortunate part of American political culture. A "Republican Form of Government" requires of its citizens, among many other duties, constant willingness to ask if we in fact have such a government and, if not, how the Constitution might in fact be improved. After all, as James Madison himself wrote in *Federalist No. 14*, my own favorite among the eighty-five essays comprising *The Federalist*, the framers in Philadelphia "formed the design of a great Confederacy, which it

---

* W. St. John Garwood and W. St. John Garwood, Jr. Centennial Chair in Law, University of Texas Law School; Professor of Government, University of Texas at Austin.

1. Kathleen M. Sullivan, *Constitutional Amendmentitis*, THE AMERICAN PROSPECT (Sept. 21, 1995), http://prospect.org/cs/articles?article=constitutional_amendmentitis.

2. *See* SANFORD LEVINSON, OUR UNDEMOCRATIC CONSTITUTION: WHERE THE CONSTITUTION GOES WRONG (AND HOW WE THE PEOPLE CAN CORRECT IT) 176–77 (2006).

3. *See* Sullivan, *supra* note 1.

4. *See id.*

is incumbent on their successors to improve and perpetuate."[5] I interpret this as suggesting that perpetuation will *require* improvement rather than complacent acceptance of any given constitutional status quo.

No one participating in this symposium, whatever his or her politics, would wish to defend the Constitution of 1787 as a perfect document, even as amended afterward with the addition of the Bill of Rights. All, I am confident, believe that the Reconstruction Amendments were necessary to overcome what was obviously the greatest imperfection. And the willingness to suggest any amendments today, whether one's politics are tilted to the right or left, manifests a belief that our constitutional order could in fact be made better. All of us agree on that abstract point, even if we disagree, perhaps vehemently, about what would constitute improvements.

I was glad to see that Professor Reynolds quoted from Marshall's opinion in *McCulloch v. Maryland*, which I believe is possibly the most important single opinion in our history. For years, I confess, I found somewhat mysterious Felix Frankfurter's remarkably effusive statement that the reminder that "'we must never forget, that it is *a constitution* we are expounding'" was "the single most important utterance in the literature of constitutional law—most important because most comprehensive and comprehending."[6] You might ask yourselves what is so amazing about this sentence, since it is glaringly obvious that Marshall and his colleagues were doing just that. What makes the sentence worthy of Frankfurter's approbation, though, is what follows—and is missing from Professor Reynolds' essay—where Marshall, in effect, sets out what is most important about legal documents that we call "constitutions."[7] Thus, he emphasizes that the United States Constitution is "intended to endure for ages to come, and, consequently, to be adapted to the various *crises* of human affairs."[8] Interestingly enough, he italicizes "crises," though one might believe that the term "to be adapted" is equally worthy of emphasis.

The point is that John Marshall—sometimes apotheosized as "the Great Chief Justice"[9]—recognized that the United States Constitution *must* be a "living Constitution" (a term that, of course, he did not use) if it was to achieve the most fundamental purpose of "endur[ing] for ages to come."[10] In this belief, he was a faithful disciple of Madison and even Thomas

---

5. THE FEDERALIST NO. 14 (James Madison).

6. Felix Frankfurter, *John Marshall and the Judicial Function*, 69 HARV. L. REV. 217 (1955) (quoting McCulloch v. Maryland, 17 U.S. 316, 407 (1819)), *reprinted in* OF LAW AND MEN: PAPERS AND ADDRESSES OF FELIX FRANKFURTER 1939–1956 5 (Philip Elman ed., 1956).

7. *See McCulloch*, 17 U.S. at 407, 415.

8. *Id.* at 415.

9. *See, e.g.*, CHARLES F. HOBSON, THE GREAT CHIEF JUSTICE: JOHN MARSHALL AND THE RULE OF LAW 214 (1996).

10. *McCulloch*, 17 U.S. at 415.

Jefferson, Marshall's despised adversary, who wrote that just as "manners and opinions change with the change of circumstances, institutions must advance also, and keep pace with the times."[11]

One obvious form of adaptation is amendment; another, however, is latitudinarian interpretation of the document—especially in response to the actions taken by other branches of government—that changes, sometimes in quite fundamental respects, our expectations about what is possible (or, less often, impossible) from government. One might agree that explicit "amendatory adaptation" is better than "amendment by latitudinarian interpretation," but, as Mr. Lynch especially recognizes, a significant deficiency of the United States Constitution is that it is quite literally the most difficult to amend constitution in the entire world.[12] To rely on formal amendment to get needed "adaptations" is to condemn ourselves to futility and to the consequences of what might indeed be highly destructive features of the "unadapted" constitutional order. Should there be any doubt, I am glad to specify my agreement with his argument that perhaps the most valuable single amendment would be to make the process of amendment significantly easier than it is now. I regret that he did not present a more specific proposal that might further this important conversation. One thing that law professors might do is to pay some attention to the fifty state constitutions that constitute an important part of the American constitutional tradition; instead, law professors tend to ignore what can be learned from looking at these constitutions because of excessive concentration on the quite idiosyncratic national Constitution.[13] One will find a plethora of amendatory procedures, each and every one of them providing an easier path to amendment than did the drafters of the 1787 Constitution that continues to structure our national polity.

Many contemporary Tea Partiers have become almost fanatical in their denunciation of "progressive constitutionalism," by which they mean the "evolutionary" approach to American constitutionalism enunciated by, among others, President Woodrow Wilson. But Wilson comes along quite late in the game. Marshall, like Madison or Jefferson, obviously could not have read the writings of Charles Darwin, but one suspects that he, like Wilson, would have been sympathetic to Darwin's basic insights about the consequences of mulish refusal to adapt—death. That was, after all, the lesson taught by the Philadelphia Convention, which engaged in spectacular disregard of its limited mandate from Congress or the requirements of Article XIII of the Articles of Confederation regarding amendment in order

---

11. Letter from Thomas Jefferson to Samuel Kercheval (July 12, 1816), *quoted in* LEVINSON, *supra* note 2, at xi.

12. *See* Timothy Lynch, *Amending Article V to Make the Constitutional Amendment Process Itself Less Onerous*, 78 TENN. L. REV. 823, 829 (2011).

13. *See generally* JOHN DINAN, THE AMERICAN STATE CONSTITUTIONAL TRADITION (2006).

to provide a new political order adequate to the exigencies facing the new country.

Given the fondness for federalism expressed by several contributors to this symposium, it is worth noting that one of the quite valuable lessons we can learn from many states is the desirability not only of easier processes of amendment, but also of scrapping constitutions that are perceived as no longer efficacious in favor of more up-to-date ones. Thus I note that Tennessee has had three constitutions, while Florida, Professor Foley's home state, has had six such documents, the most recent adopted in 1969.[14] Perhaps this constitutes the kind of change we can all believe in! In any event, everyone in this symposium accepts the desirability of adaptive change, even if, by the rules of the symposium, their arguments take the form of suggesting new amendments (none of which, for better or worse, thanks to the truly egregious Article V, has the slightest chance of being either proposed or ratified, or both).

Several of the proposals are nostalgic for a past when the national government "knew its place" relative to the powers of what Marshall, sincerely or not, labeled the "sovereign state[s]" in *McCulloch* itself. This is most obvious in Professor Foley's proposal, which comes, after all, in an essay entitled *Sovereignty, Rebalanced: The Tea Party and Constitutional Amendments*,[15] but it surely pervades the contributions of Professors Reynolds, Barnett, and Lynch.

As already suggested, I am almost wholly unsympathetic with their respective proposals. For those interested in my particular critique of Professor Barnett's "Repeal Amendment," which I have publicly described as a "terrible idea," I refer readers to a symposium published on the *Northwestern Law Review* website, based on an American Association of Law Schools panel that included, among others, Professor Barnett and myself, on "Tea Party Constitutionalism."[16] But to appreciate the specifics of my critique, one should be more accurate in describing my position than Professor Foley is. Thus, she implies that my opposition to the Repeal Amendment is equivalent to opposing federalism per se. Unfortunately for her argument, that is not what I said nor, indeed, is it a fair interpretation of

---

14. *See* Donald Lutz, *Toward a Theory of Constitutional Amendment*, *in* RESPONDING TO IMPERFECTION: THE THEORY AND PRACTICE OF CONSTITUTIONAL AMENDMENT 248–49 (Sanford Levinson ed., 1995).

15. Elizabeth Price Foley, *Sovereignty, Rebalanced: The Tea Party & Constitutional Amendments*, 78 TENN. L. REV. 751 (2011).

16. *See generally* NW. U. L. REV. COLLOQUY, http://colloquy.law.northwestern.edu/ (specifically, see the entries spanning Mar. 27, 2011 through May 16, 2011); Sanford Levinson, *If We Have an Imperfect Constitution, Should We Settle for Remarkably Timid Reform? Reflections Generated by the General Phenomenon of 'Tea Party Constitutionalism' and Randy Barnett's Particular Proposal for a 'Repeal Amendment' Designed to Rein in an Overreaching Congress*, 105 NW. U. L. REV. COLLOQUY 271 (2011), http://www.law.northwestern.edu/lawreview/colloquy/2011/9/LRColl2011n9Levinson.pdf.

my argument. She does accurately quote my belief that the Barnett proposal is a "'terrible idea'" because it would give "'outsize influence'" to "'small parochial rural states in which most Americans do not live.'"[17] She then offers her own gloss on this comment, which is the unwarranted assertion that "[t]he fact that Levinson doesn't like giving equal weight to small and large states tells us that he's not a big fan of the concept of federalism."[18] That is a pure non sequitur.

Federalism, as she herself recognizes in the title of her essay, involves a decision to assign meaningful attributes of "sovereignty" to sub-national units. I basically agree with Malcolm Feeley and Ed Rubin that it is not a very good idea,[19] so she *is* correct that I am "not a big fan of the concept of federalism."[20] But the central point is that *truly strong federalism would be a bad idea if every state in the Union were of precisely equal population.* My views of federalism have literally nothing to do with the fact that the peculiar American variety of that practice gives unwarranted power to small states. I also agree with Feeley and Rubin that adopting decentralization as a policy is often a very good idea; it is indeed quite naïve to believe that all important policies should be created or implemented from the center. Indeed, even France and China, as Jenna Bednar notes in her recent defense of federalism,[21] have sensibly adopted many decentralized public policies. Doubts about the merits of constitutionally guaranteed state autonomy do not in the least entail that one necessarily prefers centralization, just as doubts about the wisdom of entrenching the power of remarkably small states do not at all require that one reject federalism itself.

Professor Foley might be surprised that, in the San Francisco panel mentioned above, I sincerely suggested that I might be willing to sign on to the Repeal Amendment *if*, and only if, my friend Randy Barnett modified it to require that, for example, a majority of states with a majority of the American population agree that a federal statute deserved repeal. Note that I wouldn't even necessarily require a supermajority of states as Barnett does; a majority would be enough, as long as they contained a majority of our fellow Americans. Interestingly enough, when defending his unamended Amendment, Professor Barnett was reduced to arguing that it was exceedingly unlikely that the two-thirds of the states that would come together to demand repeal would in fact not contain a majority of the population.[22] Perhaps he is correct, but, if so, then it should be a relatively

17. *See* Foley, *supra* note 15, at 762 (quoting Kate Zernike, *Proposed Amendment Would Enable States to Repeal Federal Law*, N.Y. TIMES, Dec. 19, 2010, at A14.).

18. *Id.*

19. MALCOLM M. FEELEY & EDWARD RUBIN, FEDERALISM: POLITICAL IDENTITY AND TRAGIC COMPROMISE 150–53 (2008).

20. Foley, *supra* note 15, at 762.

21. *See* JENNA BEDNAR, The ROBUST FEDERATION 22 (2009).

22. *See* Randy Barnett, *The Case for the Repeal Amendment*, 78 TENN. L. REV. 813, 820–21 (2011).

minor concession to "democrats" (note the lack of capitalization) like myself who quite literally hate the further entrenchment of even the possibility of minority rule by small states in the American constitutional order.

Perhaps it is equivalent to the Devil quoting scripture, but I am quite happy once more to invoke James Madison as authority for my view. After all, in *Federalist No. 62*, he described the equal allocation of power to small and large states alike in the Senate as a "lesser evil."[23] What justified this evil was the raw fact that submission to Delaware's extortionate demand was an empirical condition precedent to achieving the desired replacement of the dysfunctional Articles of Confederation, given the exigencies of 1787 politics and the geographical location of Delaware. He was correct in his description of what is often called the "Great Compromise." But that Great Compromise was only marginally better than the other Great Compromise that gave us the Constitution, which was pandering to the interests of slave states by, most importantly, giving them added representation in the House of Representatives and, therefore, the Electoral College, through the Three-Fifths Clause.

I am confident that Professor Foley would not recommend returning to *this* aspect of 1787 Constitutionalism, even if, after all, it was a means of protecting the most important reality of American federalism, the existence of chattel slavery. I believe that one should be no more eager to defend further entrenchment of small-state power, unless one agrees with the less attractive legacy of Thomas Jefferson that people who live in rural areas are simply better than those of us (including Professor Foley) who have the misfortune to live in cities in basically urban states like Texas or Florida. She writes that my

> use of the pejorative adjective "parochial" to describe small states reveals a common liberal bias against rural America, which liberals fault for clinging too tightly to guns, Bibles and the Constitution. It is much better, under this elitist liberal view, to let densely populated, 'sophisticated' urban areas dominate the legal system.[24]

Perhaps I was undiplomatic in using the word "parochial," though I think it can easily be demonstrated that an almost logical entailment of living in small-population states is that one will have fewer opportunities to appreciate (or, for that matter, reject) in one's everyday life the remarkable diversity of contemporary American life. I presume that (particularly) southern Florida is quite distinctively different even from northern Florida, let alone, say, Wyoming or Montana. (How many Cuban restaurants does the Upper Midwest have?)

---

23. THE FEDERALIST NO. 62 (James Madison).

24. Foley, *supra* note 15 at 762.

Is it relevant that perhaps my favorite state in the present Union is the distinctly rural Vermont, the home of Ben and Jerry's that is represented in the Senate by an avowed democratic socialist, Bernie Sanders? But that still does not justify describing Vermont as particularly "diverse" (in comparison to, say, Texas or Florida) or, more to the point, granting the Green Mountain State equal representation with, say, Texas, even if I am decidedly no fan of either of my current senators. Nor does it justify the abhorrent assignment of equal voting power to the states should the House of Representatives be tasked with breaking a deadlock in the Electoral College—something, I have found, that not even vigorous defenders of the Electoral College—as against popular election—are willing to defend. Forget even popular election, which I support. Would Professor Foley oppose a constitutional amendment that would simply allow the House to pick a President by a majority vote of its total membership? Would she describe *that* as manifesting indefensible hostility to federalism?

I can easily concede that people who live in small states are uniformly fine and upstanding in all respects without having to agree that they should be the recipients of what can only be described as a spectacular "affirmative action" program going almost infinitely beyond simply putting a thumb on the scale when making admissions or hiring decisions at a law school or university. Not even those of us who support traditional affirmative action programs support giving members of our favorite groups extra votes at election time. But she presumably sees nothing at all problematic in giving Wyomingites, say, the equivalent of seventy-two votes, as compared with every voter in California; or giving the roughly one-third of the American population who live in the thirty-four least-populated states the theoretical ability to veto legislation passed by both Houses of Congress and, in most cases, signed by the President of the United States, were the "Repeal Amendment" actually part of the Constitution (which it will never be).

The United States purports to be a democracy, at least in its public presentation, and American Presidents from at least Woodrow Wilson to George W. Bush and Barack Obama, have declared that almost the essence of the American political project is to "make the world safe for democracy" or to spread the blessings of democracy abroad.[25] Consider in the effusive praise offered by President George W. Bush for Natan Sharansky's *The Case for Democracy: The Power of Freedom to Overcome Tyranny and Terror*. Indeed, a columnist for *The Economist* described President Bush as having an "intellectual love affair" with the Israeli author and politician.[26] "'I want you to read a book,'" he told John Dickerson of *Time* magazine.[27]

---

25. *See The Pope and the War*, 116 THE OUTLOOK 642, 643 (1917).

26. Lexington, *The Odd Couple: George Bush is Having an Intellectual Love Affair*, THE ECONOMIST, Feb. 3, 2005, at 32, *available at* http://www.economist.com/node/3623386?Story_ID=3623386.

27. *Id.*

"'It will give you a sense of what I'm talking about.'"[28] The *Economist* columnist "Lexington" notes that after Sharansky had run into Dick Cheney "at an American Enterprise Institute retreat in Beaver Creek, Colorado in June 2002, and had a long conversation with him," the President only a few days later "gave a speech telling the Palestinians that they needed to embrace democracy."[29] One wonders if professors Barnett and Foley would agree, at least if by democracy one means some due regard for "majority rule."[30]

Indeed, it is their disdain for majority rule that makes the seemingly radical proposals endorsed by Barnett and Foley almost truly pointless. As Barnett especially concedes,[31] it is spectacularly unlikely that the Repeal Amendment would in fact be consequential, given the difficulty of putting together a coalition of a minimum of sixty-seven state legislative houses (assuming that one of the states is the unicameral Nebraska) in thirty-four states to override some offensive federal statute. If one truly wanted to shake things up, then why not suggest emulating American states like Maine and California, both of which offer their citizens the ability to override ostensibly offensive legislation by initiative and referendum? Or, if one believes that even Republican-dominated Supreme Courts are unwilling to do very much to restore genuinely meaningful "state sovereignty," then why not propose a constitutional amendment stating that, say, five of the nine members of the Supreme Court be appointed by the states? Perhaps one could take a leaf from the very interesting Article VI, Section 2 of Tennessee's Constitution, which provides that "The Supreme Court shall consist of five judges, of whom not more than two shall reside in any one of the grand divisions of the state."[32] So the five state-selected judges might come from appropriate "grand divisions" of the United States, with the nominees selected by, say, the collective decision of the governors of the states in a given division or by the collective Chief Justices of the state supreme courts (none of whom, of course, could be eligible for selection).

Given the degree of outrage expressed at the purported depredations of the national government, it is quite remarkable how timid most of the proposals are. Perhaps I should make an exception for Professor Reynolds' suggestion that indigents, in effect, be deprived of their right to vote (unless they happen to pay federal income tax, which is presumably unlikely).[33]

---

28. *Id.*

29. *Id.*

30. Consider, for starters, what happened in Gaza when free elections were held, though, as a matter of fact, Hamas prevailed because the Palestinian Authority was too naïve to figure out the implications of the electoral rules that structured the election. *See* James Glanz, *A Lesson From Hamas: Read the Voting Law's Fine Print*, N.Y. TIMES, Feb. 19, 2006, at C4, *available at* http://www.nytimes.com/2006/02/19/weekinreview/19glanz.html.

31. *See* Barnett, *supra* note 22, at 817.

32. TENN. CONST. art. VI, § 2.

33. *See* Glenn Harlan Reynolds, *Foreword: Divine Operating System?*, 78 TENN. L.

Here, too, there is a nostalgic hearkening back to the distinctly undemocratic views of many of the framers, who believed that only those with property could be trusted to make decisions in the public interest, instead of the poor rabble that would undoubtedly use their political power, if they had it, to redistribute resources from the "haves" to themselves. This fear, of course, is what is behind the defense of strong judicial review in *Federalist No. 78*, where what we might today call the "discrete and insular minority" that Hamilton is worried about involves the propertied classes who are threatened by democracy.[34]

Hamilton was, of course, correct. Neither the Contract Clause nor the Takings Clause has done much to protect the rights of the propertied. The current Supreme Court is trying its best to help out with such decisions as *Citizens United*, which offers corporate leaders the ability to spend vast sums in the political process,[35] but, of course, it's not at all clear if that decision, whatever one thinks of it, will really be particularly consequential. Far more efficacious would be depriving the unpropertied of a right to vote, even if it is hard to imagine that existing political officials seeking re-election would be eager to go to the hustings and explain that they were trying to disenfranchise significant numbers of their present constituents.[36]

Professor Reynolds, to his credit, recognizes the relevance of very basic questions of constitutional design to political outcomes, including perhaps the most basic question: Who is entitled to participate in making decisions? What I admire about some members of the Tea Party, whether I agree with them or not, is their recognition of that basic point. If one believes that the present Senate is insufficiently protective of state sovereignty, as it surely is, then one plausible response is to suggest repeal of the Seventeenth Amendment and return of the selection of senators to state legislatures. That would presumably create an incentive on the part of senators seeking re-election to be attentive to the interests of state political officials. Better, of course, would be providing those officials with the ability to "recall" senators whose votes indicate insufficient commitment to the state's conception of its own interest. Even better would be adopting a page from the German constitution, in which the members of the Bundesrat are active state officials.[37]

REV. 651, 656–57 (2011).

34. *See* United States v. Carolene Prods. Co., 304 U.S. 144 n.4 (1937); *see generally* THE FEDERALIST NO. 78 (Alexander Hamilton) (describing the judicial branch of the federal government and its importance in defending property rights).

35. *See* Citizens United v. Fed. Election Comm'n, No. 08-205, slip op. at 50 (U.S. Jan. 21, 2010).

36. No doubt the passage of strong voter-ID laws is felt by some to accomplish some of the same ends, though, obviously, never defended in such terms.

37. *See* GRUNDGESETZ FÜR DIE BUNDESREPUBLIK DEUTSCHLAND [GRUNDGESETZ] [GG] [BASIC LAW], May 8, 1949, BGBl. I at 371, art. 50–51 (Ger.).

Or, if one is truly serious about restoring "state sovereignty," then we might address whether the United States Constitution might emulate the now-discarded Soviet Constitution, Article 72 of which specified the right of each constituent Republic of the Union of Soviet Socialist Republics to "freely secede" from the USSR.[38] Perhaps a less loaded analogue is the Lisbon Treaty that now binds the European Union, which also includes a clause specifying how disgruntled members might withdraw from the Union.[39] Southern arguments for a right to secession in 1860–61 were scarcely stupid,[40] even if the particular cause for which it was invoked was evil. We are assured by contemporary devotees of federalism that their support no longer has anything to do with racial discrimination. So why not return to the great debates about the propriety of "exit" as the ultimate response to one's alienation from organizations that are viewed as no longer sufficiently attentive to one's deepest interests (or "rights")?[41] For all of the blustering about an overreaching national government, is it possible that most of the blusterers are in fact loyal Americans who believe that in fact we (and they) are better off in the Union that exists, whatever its flaws, than they would be in a new "sovereign state," whether we call it Pacifica, New England, or Dixie? (Texas and California are presumably the only states that could plausibly make a go of it on their own.)

---

38. KONSTITUTSIIA SSSR (1977) [KONST. USSR] [USSR CONSTITUTION] art. 72 (Russ.), *available at* http://www.departments.bucknell.edu/russian/const/77cons03.html (last visited May 5, 2011).

39. Treaty of Lisbon Amending the Treaty on European Union and the Treaty Establishing the European Community, Dec. 17, 2007, 2007 O.J. (C306) 40–41. Article 49A:

> 1. Any Member State may decide to withdraw from the Union in accordance with its own constitutional requirements.
> 2. A Member State which decides to withdraw shall notify the European Council of its intention. In the light of the guidelines provided by the European Council, the Union shall negotiate and conclude an agreement with that State, setting out the arrangements for its withdrawal, taking account of the framework for its future relationship with the Union. That agreement shall be negotiated in accordance with Article 188 N(3) of the Treaty on the Functioning of the European Union. It shall be concluded on behalf of the Union by the Council, acting by a qualified majority, after obtaining the consent of the European Parliament.
> 3. The Treaties shall cease to apply to the State in question from the date of entry into force of the withdrawal agreement or, failing that, two years after the notification referred to in paragraph 2, unless the European Council, in agreement with the Member State concerned, unanimously decides to extend this period.

Id.

40. *See* Sanford Levinson, *"Perpetual Union," "Free Love," and Secession: On the Limits to the "Consent of the Governed"*, 39 TULSA L. REV. 457 (2004).

41. *See* ALBERT HIRSCHMAN, EXIT, VOICE, AND LOYALTY 112–14 (1970).

My hope is that this symposium will be the first of many in which serious and thoughtful people of various political persuasions will address what they believe to be defects in the contemporary American constitutional order and what kinds of amendatory changes might be "necessary and proper" to change it in desirable directions. It should be obvious that much more could be said, not only about the particular proposals offered us, but about the reasons why one might wish a significantly changed Constitution and what explicit forms those changes should take.

TENNESSEE LAW REVIEW
Member of the National Conference of Law Reviews

Volume 78 | Summer 2011 | Number 4

CONTENTS

ARTICLES

PROCEED AT YOUR PERIL: CROWDFUNDING AND THE SECURITIES ACT OF 1933
Joan MacLeod Heminway & Shelden Ryan Hoffman 879

SALES TO GRANTOR TRUSTS: A CASE STUDY OF WHAT THE IRS AND CONGRESS CAN DO TO CURB AGGRESSIVE TRANSFER TAX TECHNIQUES
Jay A. Soled & Mitchell Gans 973

PROPERTY RIGHTS, PROPERTY WRONGS, AND CHATTEL DISPOSSESSION UNDER SELF-STORAGE LEASES
Jeffrey Douglas Jones 1015

JUSTIFYING GINA
Abigail Lauren Perdue 1051

NO WALK IN THE DOG PARK: DRAFTING ANIMAL CRUELTY STATUTES TO RESOLVE DOUBLE JEOPARDY CONCERNS AND ELIMINATE UNFETTERED PROSECUTORIAL DISCRETION
Laurie Serafino 1119

ANNUAL INDEX 1171

PROCEED AT YOUR PERIL: CROWDFUNDING AND THE SECURITIES ACT OF 1933

JOAN MACLEOD HEMINWAY* & SHELDEN RYAN HOFFMAN**

I. INTRODUCTION 880
II. ARE CROWDFUNDING INTERESTS SECURITIES UNDER THE SECURITIES ACT? 885
A. Howey *Explained and Interpreted* 886
B. Howey *Applied in Context* 890
1. Contract, Transaction, or Scheme 892
2. Investment of Money 895
3. Common Enterprise 901
4. Expectation of Profits 902
5. Solely from the Efforts of Others 903
C. The Question of Context 904
III. FEDERAL REGULATION OF OFFERINGS OF CROWDFUNDING INTERESTS AS SECURITIES 906
A. Applicable Regulation 907
1. Registration and Related Liability and Costs 907
2. Exemptions from Registration 911
a. Private Offering Exemption under Section 4(2) of the Securities Act 912
b. Rules 504, 505, and 506 of Regulation D 916
c. Regulation A 921
3. Whose Conduct is Regulated? 922

---

* College of Law Distinguished Professor of Law, The University of Tennessee College of Law; J.D., New York University School of Law; A.B., Brown University. Work on this article has been supported by research funding from The University of Tennessee College of Law and has benefited significantly from the capable research and writing of my coauthor, S. Ryan Hoffman and the inspiration and encouragement of Josh Tetrick. I first began to study crowdfunding ventures in the spring of 2009, and I shared many of the ideas presented here at faculty forums at the University of Florida Fredric G. Levin College of Law, Mississippi College School of Law, and The University of Tennessee Corporate Governance Center and academic presentations at the 2011 Annual Meeting of the Law and Society Association, the 2011 Annual Conference of the Southeastern Association of Law Schools, and CLEA (Canadian Law and Economics Association) 2011. I am indebted to each of those audiences for their input on this article.

** M.B.A. candidate, The University of Tennessee College of Business Administration; J.D., The University of Tennessee College of Law; A.B., Clemson University. I would like to thank Professor Heminway for including me in this project in which she had invested considerable time before inviting me to participate. I am grateful to her for sharing her knowledge and insight and for providing the invaluable guidance that made this article possible.

B. Policy Underpinnings 927
IV. EXEMPTING CROWDFUNDING OFFERINGS FROM SECURITIES ACT REGISTRATION REQUIREMENTS 930
A. Why Consider an Exemption for Crowdfunding? 930
1. Perceived and Actual Advantages of Crowdfunding..... 931
2. Perceived and Actual Disadvantages of Crowdfunding.... 933
B. The Contours of a Possible Crowdfunding Exemption 937
1. Foundational Principles..... 937
2. Rule-Making Options and Processes..... 941
3. Substantive Elements of a Proposed Exemption.....945
V. CONCLUSION 961

I. INTRODUCTION

Funding small businesses while complying with applicable securities laws and regulations is tricky. Businesses need money to commence and sustain operations, but most non-bank funders desire to subscribe for interests in the firm that are deemed securities, invoking the possibility that federal or state securities laws regulate the transaction. Under federal law, an offer or sale of securities must be registered unless the security or the transaction is exempt.[1]

For many small businesses, the cost of complying with applicable regulatory requirements outweighs the benefits associated with the proposed financing method. Small firms—even those with good ideas and sustainable business plans—either never get their fair shake in obtaining start-up funds, or fail because they cannot finance the continued operation or growth of the business. There must be a better way.

A promising web-based funding model for small business firms (and potentially for larger enterprises) has emerged over the past few years. Crowdfunding, as this model has come to be known, has been defined and described in various ways.[2] A recent article explains crowdfunding as follows:

---

1. *See* 15 U.S.C. § 77e (2006).

2. *See, e.g.*, JEFF HOWE, CROWDSOURCING: WHY THE POWER OF THE CROWD IS Driving Business 281 (2008) [hereinafter HOWE, CROWDSOURCING] ("Crowdfunding taps the collective pocketbook, allowing large groups of people to replace banks and other institutions as a source of funds."); Kristina Dell, *Crowdfunding*, TIME.COM (Sept. 4, 2008), http://www.time.com/time/magazine/article/0,9171,1838768,00.html (describing crowdfunding as "[o]ne part social networking and one part capital accumulation"). Some define crowdfunding to exclude websites that promise to pay profit sharing to the funders. *See* Kieran Masterton, *How to Crowdfund Your Startup*, THINK VITAMIN (Apr. 22, 2010), http://thinkvitamin.com/web-industry/how-to-crowdfund-your-startup/ ("[R]ule number one of Crowdfunding is that you don't offer a percentage of your venture as a reward."). Crowdfunding is sometimes confused or conflated with crowdsourcing, which also is a

> The concept of crowdfunding finds its root in the broader concept of crowdsourcing, which uses the "crowd" to obtain ideas, feedback and solutions in order to develop corporate activities. In the case of crowdfunding, the objective is to collect money for investment; this is generally done by using social networks, in particular through the Internet (Twitter, Facebook, LinkedIn and different other specialized blogs). The crowd-funders (those who provide the money) can at times also participate in strategic decisions or even have voting right. In other words, instead of raising the money from a very small group of sophisticated investors, the idea of crowdfunding is to obtain it from a large audience (the "crowd"), where each individual will provide a very small amount.[3]

Crowdfunding includes a variety of business financing models that use the Internet.[4] As we use the term in this article, crowdfunding involves using a web-based business enterprise to seek and obtain incremental venture funds from the public using a website (which we refer to as a "crowdfunding website") to connect businesses or projects in need of funding (which we refer to as "crowdfunded ventures") with potential funders.[5] While some crowdfunding websites specialize in a particular crowdfunded venture or type of crowdfunded venture, others are more

---

poorly defined term. Jeff Howe, the apparent originator of the term, defines it broadly to include collaborative online business ventures of many kinds, where the power of the Internet is harnessed to substitute the public for employees, creating a new type of labor market. *See* Jeff Howe, *The Rise of Crowdsourcing*, WIRED (June 2006), http://www.wired.com/wired/archive/14.06/crowds.html [hereinafter Howe, *Rise of Crowdfunding*]. One commentator connects the two terms by offering that "[t]he term crowdfunding derives from another neologism: crowdsourcing, i.e., outsourcing to the public jobs typically performed by employees." Dell, *supra*. Yet another includes a crowdfunding website, Aswarmofangels.com, in his list of crowdsourcing "efforts," implying that crowdfunding is a type of crowdsourcing. Jessi Hempel, *Tapping the Wisdom of the Crowd*, BLOOMBERG BUSINESSWEEK (Jan. 18, 2007, 11:01 AM), http://www.businessweek.com /innovate/content/jan2007/id20070118_768179.htm. Jeff Howe agrees, noting that "crowdfunding has more in common with other forms of crowdsourcing than is immediately apparent." HOWE, CROWDSOURCING, *supra*, at 247.

3. Paul Belleflamme et al., Crowdfunding: Tapping the Right Crowd 2 (Feb. 21, 2011) (unpublished manuscript), *available at* http://ssrn.com/abstract=1578175.

4. *See* KEVIN LAWTON & DAN MAROM, THE CROWDFUNDING REVOLUTION 1 (2010) ("[T]he true social vibrance of crowdfunding is difficult to capture in a definition. The crowdfunding space is quite diverse, comprised of many niches, and shares a lot of social networking's energy."); C. Steven Bradford, Crowdfunding and the Federal Securities Laws 9-17 (Oct. 7, 2011), http://ssrn.com/abstract=1916184 (accepted for publication in the Columbia Business Law Review) (describing various different models).

5. This definition is not vastly different from the one constructed by the authors of the article cited *supra* note 3. These authors define crowdfunding, based on the fifty-one examples they reviewed, as involving "an open call, essentially through the Internet, for the provision of financial resources either in form of donation or in exchange for some form of reward and/or voting rights in order to support initiatives for specific purposes." Belleflamme et al., *supra* note 3, at 6.

general. A chart identifying and describing a number of crowdfunding websites is included as Table 1.[6]

We became interested in this venture finance model because it has huge appeal in a number of obvious respects, yet we could not understand how some of the crowdfunding websites and crowdfunded ventures (especially those offering profit-sharing interests to funders) were complying with federal securities laws.[7] This article is the result of our study of these firms in that context and includes both descriptive and normative observations. It is clear that some but not all manifestations of the crowdfunding model result in the offer and sale of interests that are securities under the Securities Act of 1933, as amended (the "Securities Act").[8] Because the offers and sales are neither registered nor exempt from registration as required under the Securities Act, these crowdfunded ventures are at risk. They are the focus of the analysis and prescriptions we offer here.

We are not alone in our engagement with the federal securities law aspects of crowdfunding. Among others, the U.S. Securities and Exchange Commission ("SEC") and the U.S. Congress have taken an interest in crowdfunding. In an April 2011 letter to the Chairman of the Committee on Oversight and Government Reform of the U.S. House of Representatives, SEC Chairman, Mary Schapiro expressly mentions crowdfunding as a new capital-raising strategy.[9] She further indicates that the SEC has been

---

6. Table 1 features websites in existence as of the date we completed this article, unless otherwise expressly noted.

7. We focus in this article on the federal laws and regulations governing securities offerings—most specifically the Securities Act and rules that the SEC adopted under it. We note here, however, the potential applicability of aspects of broker-dealer, investment advisory, or exchange regulation and state securities (or "blue sky") laws and regulations. *See* Bradford, *supra* note 4, at 30–51; Thomas Lee Hazen, Crowdfunding, Social Networks, and the Securities Laws—The Inadvisability of a Specially Tailored Exemption Without Imposing Affirmative Disclosure Requirements 6, 15–16 (Nov. 18, 2011) (unpublished manuscript), *available at* http://ssrn.com/abstract=1954040; C. Steven Bradford, *Peer-to-Peer Lending, Crowd-funding, and Securities Law*, BUS. L. PROF. BLOG (June 17, 2011), http://lawprofessors.typepad.com/business_law/2011/06/peer-to-peer-lending-crowd-funding-and-securities-law.html [hereinafter Bradford, *Peer-to-Peer Lending*]. We also note that the possibilities and perils of crowdfunding cross state borders internationally. While this is extremely important, it is part of a larger issue in international (and especially Internet) securities offerings that this article does not attempt to resolve. At various points throughout, however, as relevant to our analysis, we reference crossover issues under other federal, state, and international law. And finally, although we note in several places the possibility of other regulatory schemes (notably, the regulation of gambling and charitable solicitations), this article does not address in any meaningful way the possible application of these other potential sources of regulation.

8. 15 U.S.C. § 77a–aa (2006).

9. Letter from Mary L. Schapiro, Chairman, U.S. Sec. & Exch. Comm'n to The Honorable Darrell E. Issa, Chairman, Comm. On Oversight & Gov't Reform, U.S. House of Representatives (Apr. 6, 2011), at 22–24, http://www.sec.gov/news/press/schapiro-issa-

engaged with industry participants, and is focusing on how to address regulatory concerns relating to crowdfunding.[10] Later, in testimony before that House Committee, Chairman Schapiro echoed many of the same themes from the letter, without expressly mentioning crowdfunding:[11]

> I recently asked the staff to take a fresh look at our offering rules in light of changes in the operation of the markets, advances in technology and the acceleration in the pace of communications. I also requested that the staff think creatively about what the SEC can do to encourage capital formation, particularly for small businesses, while maintaining important investor protections. Areas of focus for the staff will include:
>
> - the restrictions on communications in initial public offerings;
> - whether the general solicitation ban should be revisited in light of current technologies, capital-raising trends and our mandates to protect investors and facilitate capital formation;
> - the number of shareholders that trigger public reporting, including questions surrounding the use of special purpose vehicles that hold securities of a private company for groups of investors; and
> - regulatory questions posed by new capital raising strategies.
>
> In conducting this review, we will solicit input and data from multiple sources, including small businesses, investor groups and the public-at-large. The review will include evaluating the recommendations of our annual SEC Government-Business Forum on Small Business Capital Formation, as well as suggestions we receive through an e-mail box we recently created on our website. In addition, I expect our efforts to benefit from the input of the new Advisory Committee on Small and Emerging Companies the Commission is in the process of forming, which will provide a formal mechanism for the Commission to receive advice and recommendations about regulatory programs that affect privately held small businesses and small publicly traded companies.[12]

Federal legislative interest followed. First, in early November 2011, the U.S. House of Representatives passed the Entrepreneur Access to Capital Act, H.R. 2930,[13] which had been introduced in September 2011.

---

letter-040611.pdf [hereinafter Issa Letter]; *see also* C. Steven Bradford, *Crowdfunding and the SEC*, BUS. L. PROF. BLOG (Apr. 18, 2011), http://lawprofessors.typepad.com/business_law/2011/04/crowdfunding-and-the-sec.html [hereinafter Bradford, *Crowdfunding Blog*].

10. Issa Letter, *supra* note 9.

11. The Future of Capital Formation: Hearing before the H. Comm. on Oversight & Gov't Reform, 112th Cong. (2011) (testimony of Mary L. Schapiro, Chairman, Sec. & Exch. Comm'n), *available at* http://www.sec.gov/news/testimony/2011/ts051011mls.htm [hereinafter Schapiro Testimony].

12. *Id.* The email box referenced in Chairman Schapiro's testimony is available at http://www.sec.gov/spotlight/regulatoryreviewcomments.shtml.

13. H.R. 2930, 112th Cong. (2011), *available at* http://thomas.loc.gov/home/thomas.

The U.S. Senate then took up the cause. As the House passed its bill, Senator Scott Brown introduced a Senate crowdfunding bill, the Democratizing Access to Capital Act of 2011, S. 1791.[14] Hearings were held on December 1, 2011.[15] The information in this article remains current to that date, except as otherwise specified.

The federal government's current focus on crowdfunding and related capital formation strategies and business models is an important and necessary step in defining the reach of federal securities law rules in regulating crowdfunding and other innovative capital-raising methods and capital markets.[16] We assert that, with the right approach, Congress and the SEC can work with the crowdfunding industry in defining responsible parameters for crowdfunding. This article is designed to contribute to the regulatory conversation.

In our view, protecting investors and maintaining market integrity—the two principal policies underlying the federal securities laws—do not clearly compel registering interests offered and sold through crowdfunding websites, even if those interests are securities within the meaning of the Securities Act.[17] This article supports our contention in several ways. First, it explores both the foundational definitional question—whether crowdfunding interests are securities, and if so, when—and the implications of the answer to that definitional question as a matter of positive law and underlying policy. The article then addresses the advantages and disadvantages of the crowdfunding venture finance model. Finally we conclude by proposing the principles, processes, and substantive

---

php (actual bill text may be accessed by searching for appropriate bill number).

14. S. 1791, 112th Cong. (2011), *available at* http://thomas.loc.gov/home/thomas.php (actual bill text may be accessed by searching for appropriate bill number).

15. *See* Bill Summary and Status, 112th Congress (2011–2012), S. 1791, *available at* http://thomas.loc.gov/cgi-bin/bdquery/z?d112:s.01791:; *see also* U.S. Senate Committee on Banking, Housing, and Urban Affairs, *Spurring Job Growth through Capital Formation while Protecting Investors*, *available at* http://banking.senate.gov/public/index.cfm?FuseAction=Hearings.Hearing&Hearing_ID=a96c1bc1-b064-4b01-a8ad-11e86438c7e5. We note as we go to press that on December 9, 2011 a second crowdfunding bill, S. 1970, was introduced in the U.S. Senate by Senator Jeff Merkley of Oregon. *See* S. 1970, 112th Cong. (2011), *available at* http://thomas.loc.gov/home/thomas.php (actual bill text may be accessed by searching for the appropriate bill number). This article does not incorporate or address S. 1970 or any subsequent bill.

16. Another Internet-based market for small business finance that also has attracted SEC attention is the peer-to-peer lending market. *See* Angus Loten, *Peer-to-Peer Loans Grow*, WALL ST. J., June 16, 2011, at B10.

17. This question is of paramount importance to the SEC as it contemplates the appropriate level of regulation of crowdfunding. *See* Issa Letter, *supra* note 9, at 23. ("In considering whether an exemption from the registration requirements of the Securities Act is appropriate for capital formation strategies like crowdfunding, the Commission will be mindful of its dual responsibilities of facilitating capital formation and protecting investors.").

components of a registration exemption designed to enable the crowdfunding model to survive as an investment vehicle while appropriately protect those investors who cannot fend for themselves.

To accomplish these objectives, the article proceeds in four additional parts. Part II analyzes the circumstances under which crowdfunding interests are securities under the definition provided in Section 2(a)(1) of the Securities Act.[18] After situating crowdfunding interests as potential investment contracts governed by the *Howey* test (or, in the alternative, debt interests governed by the *Reves* test),[19] this Part of the article explains the elements of an investment contract and compares and contrasts them, in pertinent part, to the attributes of crowdfunding interests. This analysis reveals that some crowdfunding interests are likely classifiable as securities.[20] Given that some crowdfunding interests may be securities, Part III of the article then focuses on the consequences of that legal conclusion. Part III describes the regulatory ramifications of security status under the Securities Act's key operative provisions (which require, in significant part, that the offer and sale of a security must be registered or exempt from registration[21]) and the policies underlying both the registration requirement and relevant exemptions. Part IV explains why the offer and sale of crowdfunding interests under certain conditions should not require registration, and suggests the principles and potential parameters of a new registration exemption for crowdfunding interests, which could be adopted by the SEC under Section 3(b) of the Securities Act (although Congress would be charged with making some related changes to law to facilitate the overall exemption scheme).[22] Part V offers a summary conclusion.

II. ARE CROWDFUNDING INTERESTS SECURITIES UNDER THE SECURITIES ACT?

Our concern about the application of the Securities Act in the context of crowdfunding websites is misplaced if crowdfunding interests are not securities. Accordingly, this Part addresses the threshold question of whether crowdfunding interests are securities under the Securities Act.

The Securities Act contains a statutory definition of the term "security," which states that, "unless the context otherwise requires, the term 'security'" includes a variety of listed financial instruments.[23] The list includes, among other more typical financial interests (such as stock, bonds,

18. 15 U.S.C. § 77b(a)(1) (2006).

19. *See infra* notes 25–26, 52–55 and accompanying text.

20. *See infra* Part II.B.5.

21. *See* 15 U.S.C. § 77c–e.

22. *See id.* § 77c(b). Although we assume in this article that the SEC would initiate regulation under Section 3(b), a comparative institutional assessment to determine the appropriate rule maker should be undertaken. *See infra* note 339 and accompanying text.

23. *See id.* § 77b(a)(1).

debentures, evidence of indebtedness, and options), an "investment contract."[24] In cases involving instruments other than standard equity, debt, and derivative instruments, the application of Section 5's registration mandate often depends on whether the particular financing device is an investment contract and, therefore (unless the context otherwise requires), a security.[25] Consequently, the U.S. Supreme Court developed a common law test, known as the *Howey* test, for determining whether or not a financial instrument is an investment contract.[26] Under the *Howey* test, "an investment contract for the purposes of the Securities Act means a contract, transaction or scheme whereby a person invests his money in a common enterprise and is led to expect profits solely from the efforts of a promoter or a third party. . . ."[27] *Howey* involved the solicitation of purchasers of real property interests and related service contracts that together represented profit-sharing interests in Florida citrus groves.[28] The court found that these combined transactions constituted investment contracts and, given the context, securities.[29]

A. Howey *Explained and Interpreted*

The various parts of the *Howey* test have been illuminated in meaningful ways in subsequent federal court opinions, including a number of Supreme Court opinions. For example, in *United Housing Foundation, Inc. v. Forman*,[30] the Supreme Court better identified when an arrangement constitutes an "investment." The *Forman* Court focused on differentiating

---

24. *See id.*

25. *See, e.g.*, SEC v. Edwards, 540 U.S. 389, 392–97 (2004) (holding that a sale and leaseback arrangement of payphones on the promoter's property was an investment contract); Continental Mktg. Corp. v. SEC, 387 F.2d 466, 470–71 (10th Cir. 1967) (holding that a program for breeding beavers was an investment contract); Miller v. Cen. Chinchilla Grp., Inc., 494 F.2d 414, 417 (8th Cir. 1974) (reversing the lower court's pretrial dismissal on grounds that evidence presented at trial could possibly show that a program in which investors raised chinchillas and sold the offspring to the owners at inflated prices constituted an investment contract).

26. *See* SEC v. W.J. Howey Co., 328 U.S. 293, 298–301 (1946) (determining that an arrangement to sell profit-sharing interests in a citrus grove was an investment contract).

27. *Id.* at 298–99; *see also* United Hous. Found., Inc. v. Forman, 421 U.S. 837, 852 (1975) ("The touchstone [of an investment contract] is the presence of an investment in a common venture premised on a reasonable expectation of profits to be derived from the entrepreneurial or managerial efforts of others.").

28. The investment interests offered in the *Howey* case included a land sales contract for a citrus grove by W. J. Howey Company and an accompanying service contract with W. J. Howey Company's affiliate, Howey-in-the-Hills Service Inc. The service contract gave Howey-in-the-Hills a leasehold interest that allowed the company to cultivate the crops on the land. *Howey*, 328 U.S. at 295–96.

29. *Id.* at 300.

30. 421 U.S. 837 (1975).

"consumption" and "use" from "investment,"[31] and determined that the latter occurs when "the investor is 'attracted solely by the prospects of a return on his investment.'"[32] Thus, whether an investment exists depends on whether profits motivated the potential investor's decision to provide funds. Because the expectation of profits is an independent component of the *Howey* test,[33] we discuss profits separately below.

The U.S. Supreme Court has not defined a "common enterprise," and federal courts of appeal have taken different approaches when determining whether one exists.[34] Depending on the jurisdiction, a business venture may be a common enterprise under *Howey's* progeny if it meets one of three judicially ordained tests. Two principal types of commonality exist, one of which has two different forms.

The first type of commonality recognized in these court cases is horizontal commonality. "A horizontal common enterprise is a pool of assets . . . not separate accounts . . . ."[35] The horizontal commonality approach focuses on the relationship among the investors and requires that there be a pooling of investors' funds for the purpose of generating financial returns based on the success of the venture and, in some cases, a sharing of profits and losses on a pro-rata basis among investors.[36] For example, in *Howey*, the purchasers were offered "an opportunity to contribute money and to share in the profits of a large citrus fruit enterprise."[37] The Court noted that purchasers of the land and management

---

31. *Id.* at 852–53.

32. *Id.* at 852 (quoting *Howey*, 328 U.S. at 300). *Forman* and other federal cases typically do not focus on the word "solely" as a part of this analysis, preferring instead to look at degrees of significance. *See, e.g.*, Timmreck v. Munn, 433 F. Supp. 396, 402 (N.D. Ill. 1977) ("The court must therefore consider the nature of the promotion to determine whether the emphasis of the developers and their sales agents was on the 'investment' or the 'consumption' side of the real estate duality."). The issue in *Forman* was whether a mandatory acquisition of stock by tenants who wished to lease an apartment in a cooperative housing project involved the purchase of securities. *Forman*, 421 U.S. at 840. The Court concluded that because the tenants "purchase[d] a commodity for personal consumption or living quarters for personal use," the interests in the co-op were not securities. *Id.* at 858.

33. *Howey*, 328 U.S. at 298–99.

34. Christopher L. Borsani, *A "Common" Problem: Examining the Need for Common Ground in the "Common Enterprise" Element of the* Howey *Test*, 10 DUQ. BUS. L.J. 1, 7 (2008).

35. Schofield v. First Commodity Corp., 638 F. Supp. 4, 7 (D. Mass. 1985).

36. Borsani, *supra* note 34, at 8 (citations omitted). This approach, "as required by the Third, Sixth, and Seventh Circuits," depends on "whether [investors'] risks were pooled for a single investment purpose." *Id.* at 9 (citing Deckebach v. La Vida Charters, Inc. of Florida, 867 F.2d 278 (6th Cir. 1989)).

37. *Howey*, 328 U.S. at 299. At least one commentator, however, believes that the investment scheme in *Howey* is not characterized by a pooling of investor funds. *See* James D. Gordon III, *Defining a Common Enterprise in Investment Contracts*, 72 OHIO ST. L.J. 59, 73 (2011) ("[H]orizontal commonality was not present in *Howey* itself because each investor

contracts had "no right of entry to market the crop; thus there [was] ordinarily no right to specific fruit. The company [was] accountable only for an allocation of the net profits based upon a check made at the time of picking."[38]

Vertical commonality, the other principal type of commonality for purposes of *Howey's* investment contract definition, focuses on the relationship between the promoter and the investors, eliminating the requirement that investors pool their funds.[39] The Ninth Circuit has stated that "[a] common enterprise is one in which the fortunes of the investor are interwoven with and dependent upon the efforts and success of those seeking the investment or of third parties."[40] Thus, strict vertical commonality (as the Ninth Circuit standard has been labeled) describes a venture in which the principals or promoters do not make a profit until the investors make a profit. In other words, strict vertical commonality requires a link between investment performance and promoter remuneration.[41] The promoter must have a financial stake in the investment, and the "fortunes of the investor [are] commingled with, and dependent upon the success of the promoter."[42]

Other courts have rejected this narrow definition of a common enterprise in favor of a more open inquiry. The resultant broad vertical commonality test merely recognizes the existence of a relationship between the promoter's efforts and the investor's profits. "[T]he requisite commonality is evidenced by the fact that the fortunes of all investors are inextricably tied to the efficacy of the [promoter's efforts]."[43] Accordingly, broad vertical commonality requires a relationship between investor and promoter, but does not require that the promoter actually benefit in a manner consistent with the benefit promised to the investors (i.e., the promoter's remuneration need not be dependent upon the success of the

---

individually owned a separate tract of land."). We disagree. Although each investor purchased separate real estate, the profit-generating scheme to which each investor was contributing in that purchase involved the aggregation of their funds and lands, maintained and harvested collectively, with profits doled out from the aggregate enterprise based on the number of tracts owned. *Howey*, 328 U.S. at 295–96. In this regard, the purchaser's interests in the separate tracts of land (represented by deeds and service contracts) were the equivalent of a stock certificate in corporate equity investments—a tangible representation of each investor's profit-sharing interest. *Id.* at 300 ("Their respective shares in this enterprise are evidenced by land sales contracts and warranty deeds, which serve as a convenient method of determining the investors' allocable shares of the profits. The resulting transfer of rights in land is purely incidental.").

38. *Howey*, 328 U.S. at 296.

39. Borsani, *supra* note 34, at 9.

40. SEC v. Glenn W. Turner Enters., 474 F.2d 476, 482 n.7 (9th Cir. 1973), *cert. denied*, 414 U.S. 821 (1973).

41. Borsani, *supra* note 34, at 9–10.

42. *Id.* "[T]he Ninth Circuit is the only circuit that follows this approach." *Id* at 10.

43. SEC v. Koscot Interplanetary, Inc., 497 F.2d 473, 479 (5th Cir. 1974).

venture, but the investor's profits must be dependent upon the promoter's expertise and labor).[44] Although the circuits differ on the exact scope of the "common enterprise" requirement, horizontal commonality (which often, but not always, exists regardless of the nature of any vertical commonality in a potential investment contract arrangement) is sufficient to satisfy this prong of the *Howey* test in most federal courts.[45]

Finally, federal courts have illuminated the last two parts of the *Howey* test—the parts relating to expectations of profits from the efforts of others. The *Forman* Court defined "profits" as "either capital appreciation resulting from the development of the initial investment, . . . or a participation in earnings resulting from the use of investors' funds . . . ."[46] In a subsequent case, the Court announced that fixed returns may constitute profits under the *Howey* test.[47]

With respect to the "solely from the efforts of a promoter or a third party" prong in *Howey* (commonly shortened to "solely from the efforts of others"[48]), federal circuit court decisions have established that the proper analysis is not whether the efforts of others are the exclusive determinants of the investors' profits, but rather "whether the efforts made by those other than the investor are the undeniably significant ones, those essential managerial efforts which affect the failure or success of the enterprise."[49] Although the Court has not directly endorsed this relaxed interpretation, it "appears to have acquiesced in [this] formulation."[50] As a result, the definition of broad vertical commonality is effectively synonymous with the "efforts of others" test, conflating two of the *Howey* test prongs—the existence of a common enterprise and the generation of profits from the

---

44. Borsani, *supra* note 34, at 10 (citing SEC v. Cont'l Commodities Corp., 497 F.2d 516, 522 (5th Cir. 1974)).

45. *Id.* at 12.

46. United Hous. Found., Inc. v. Forman, 421 U.S. 837, 852 (1975).

47. SEC v. Edwards, 540 U.S. 389, 394 (2004) ("There is no reason to distinguish between promises of fixed returns and promises of variable returns for the purposes of the test, so understood.").

48. *Howey* itself used this terminology. SEC v. W.J. Howey, 328 U.S. 293, 301 (1946) ("The test is whether the scheme involves an investment of money in a common enterprise with profits to come solely from the efforts of others.").

49. SEC v. Glenn W. Turner Enters., 474 F.2d 476, 482 (9th Cir.), *cert. denied*, 414 U.S. 821 (1973); *see also* SEC v. Koscot Interplanetary, Inc., 497 F.2d 473, 480 (5th Cir. 1974) ("[A] close reading of the language employed in *Howey* and the authority upon which the Court relied suggests that . . . we need not feel compelled to follow the 'solely from the efforts of others' test literally.").

50. MARC I. STEINBERG, UNDERSTANDING SECURITIES LAWS 19 (5th ed. 2009). As dictum, the *Forman* Court acknowledged the Ninth Circuit's decision in the *Glenn W. Turner Enterprises* case, 474 F.2d at 482, which "held that 'the word 'solely' should not be read as a strict or literal limitation on the definition of an investment contract.'" *Forman*, 421 U.S. at 852 n.16. Because this issue was not presented in *Forman*, the Court expressed no view on this matter in its holding. *Id.*

efforts of others—in those jurisdictions adopting broad vertical commonality.[51]

B. Howey *Applied in Context*

Although the attributes of crowdfunding ventures differ within a range (as illustrated in Table 1), the interests that some crowdfunding websites and crowdfunded ventures offer are equity interests that afford the owners some revenue or profit-sharing rights. Our analysis in this article, and our application of the *Howey* test, both concentrate on these equity-type interests.

However, before applying *Howey* to crowdfunding interests, we note that some crowdfunding interests are styled as debt instruments. This type of instrument (separately listed under the "security" definition in Section 2 of the Securities Act as a "note, . . . bond, debenture, [or] evidence of indebtedness"[52]) likely would be analyzed under the "family resemblance" test established in *Reves v. Ernst & Young*,[53] rather than the *Howey* test. The *Reves* test begins with a "presumption that every note is a security."[54] The presumption may be rebutted by reference to a Second Circuit "list of instruments commonly denominated 'notes' that nonetheless fall without the 'security' category."[55] If, however, the instrument is not among those listed,

> [m]ore guidance . . . is needed. . . . [A]s the Second Circuit itself has noted, its list is "not graven in stone," and is therefore capable of expansion. Thus, some standards must be developed for determining when an item should be added to the list.
>
> An examination of the list itself makes clear what those standards should be. In creating its list, the Second Circuit was applying the same factors that this Court has held apply in deciding whether a transaction involves a "security." First, we examine the transaction to assess the

---

51. *See* Revak v. SEC Realty Corp., 18 F.3d 81, 88 (2d Cir. 1994) (rejecting broad commonality, and noting that under the concept, "two separate questions posed by *Howey*—whether a common enterprise exists and whether the investors' profits are to be derived solely from the efforts of others—are effectively merged into a single inquiry"); Gordon, *supra* note 37, at 76 ("Broad vertical commonality is present whenever the first, third, and fourth prongs of the *Howey* test are met—i.e., when there is an investment of money with an expectation of profits solely from the efforts of others. Thus, the broad vertical commonality test eliminates the common enterprise prong of the *Howey* test.").

52. 15 U.S.C. § 77b(a)(1) (2006).

53. *See* Reves v. Ernst & Young, 494 U.S. 56, 63 (1990) (establishing the family resemblance test in analyzing when "notes," as debt instruments listed in both the Securities Act and the Exchange Act, are securities).

54. *Reves*, 494 U.S. 56 at 65 (footnote omitted).

55. *Id.*

> motivations that would prompt a reasonable seller and buyer to enter into it. If the seller's purpose is to raise money for the general use of a business enterprise or to finance substantial investments and the buyer is interested primarily in the profit the note is expected to generate, the instrument is likely to be a "security." If the note is exchanged to facilitate the purchase and sale of a minor asset or consumer good, to correct for the seller's cash-flow difficulties, or to advance some other commercial or consumer purpose, on the other hand, the note is less sensibly described as a "security." Second, we examine the "plan of distribution" of the instrument to determine whether it is an instrument in which there is "common trading for speculation or investment." Third, we examine the reasonable expectations of the investing public: The Court will consider instruments to be "securities" on the basis of such public expectations, even where an economic analysis of the circumstances of the particular transaction might suggest that the instruments are not "securities" as used in that transaction. Finally, we examine whether some factor such as the existence of another regulatory scheme significantly reduces the risk of the instrument, thereby rendering application of the Securities Acts unnecessary.[56]

Crowdfunding interests structured in the form of interest-bearing notes or similar debt instruments are likely to be classified as securities under the *Reves* test.[57] Although crowdfunding interests do not typically trade in a secondary market, and the *Howey* test does not require an assessment of the applicability of alternative risk-reduction regulatory schemes, the same essential issues and tensions exist in the application of both the *Howey* and *Reves* tests.

Given their context, crowdfunding interests styled as equity instruments with profit-sharing components are best seen not only as investment contracts but also as securities under Section 2(a)(1) of the Securities Act[58] as interpreted by *Howey* and its progeny. Our analysis breaks the *Howey* test into five parts or prongs.[59] We add to that analysis a brief discussion of

---

56. *Id.* at 65–67 (citations omitted).

57. *See* Bradford, *supra* note 4, at 21–25 (engaging in an analysis of whether crowdfunding interests styled as debt instruments are securities and concluding that "crowdfunding notes that promise to pay interest to investors would probably be securities under the *Reves* test.").

58. 15 U.S.C. § 77b(a)(1).

59. The more standard framework for analysis separates the investment contract definition into three or four elements. *See, e.g.*, Warfield v. Alaniz, 569 F.3d 1015, 1020 (9th Cir. 2009) (applying *Howey* in three parts but acknowledging that others use a four-part test); United Am. Bank v. Gunter, 620 F.2d 1108, 1116 (5th Cir. 1980) (referring to "application of the four-pronged *Howey-Forman* test"); SEC v. Koscot Interplanetary, Inc., 497 F.2d 473, 477 (5th Cir. 1974) ("[The *Howey* test] subsumes within it three elements: first, that there is an investment of money; second, that the scheme in which an investment is made functions as a common enterprise; and third, that under the scheme, profits are derived solely from the efforts of individuals other than the investors.").

context. To illustrate our analysis, we use one of the crowdfunding websites featured in Table 1, 33needs, as our primary example.[60] Except as otherwise noted in the analysis set forth below, we assume that U.S. law applies to the purchase of the subject crowdfunding interests.[61]

1. Contract, Transaction, or Scheme

Beginning with the first prong of the *Howey* test, the language, "contract, transaction or scheme" is seemingly broad enough to cover all crowdfunding business operations. However, we have found no court decision defining these terms in this context. Jurists and legal scholars easily pass over this component of the *Howey* test and consider it satisfied because of its ostensible breadth.[62] In relevant part, the *Howey* Court notes that the overall investment contract definition it propounds originates from earlier state law opinions,[63] and that it is "immaterial whether the shares in

60. *See* 33NEEDS, http://www.33needs.com/ (last visited Sept. 9, 2011). We chose this site as our example because it allows for a more detailed treatment of the "investment of money" part of the *Howey* test. In the interest of full disclosure, it must be noted here that one of us (Professor Heminway) has been in communication with the founder of 33needs.com since January 2010. We note that, at the time final edits were made to this article, the 33needs website had been taken down in anticipation of a site redesign and redevelopment "to take advantage of the likely legislative changes re equity ownership." Email message from Josh Tetrick to Joan Heminway, Nov. 22, 2011 12.58 PM (on file with the Tennessee Law Review). The 33needs example (which was included in prior versions of the article made available on the Social Science Research Network, http://www.ssrn.com) remains a salient one, however, and we have chosen to retain it here.

61. Having assumed this, we note that the jurisdictional reach of the U.S. securities laws in this context is an unclear matter. Rule 901 of Regulation S under the Securities Act provides that "[f]or the purposes only of section 5 of the Act, the terms *offer*, *offer to sell*, *sell*, *sale*, and *offer to buy* shall be deemed to include offers and sales that occur within the United States and shall be deemed not to include offers and sales that occur outside the United States." 17 C.F.R. § 230.901 (2011) (citation omitted). Under Regulation S, an offer or sale occurs outside the United States when it is made in an "offshore transaction," which requires (among other things) that the offer not be made to a person in the United States and that any sale transaction meet other specified requirements. 17 C.F.R. § 230.903 (2011). This may not be as simple as it sounds. Moreover, despite a recent U.S. Supreme Court case, *Morrison v. Nat'l Australia Bank, Ltd.* 130 S. Ct. 2869 (2010), attempting to clarify the extraterritorial reach of the key antifraud provision applicable to purchases and sales of securities, Section 10(b) of the Securities Exchange Act of 1934, as amended, 15 U.S.C. 78j(b) (2006), questions also remain as to the application of the *Morrison* rule in specific cases.

62. *See, e.g.*, C. Edward Fletcher, III, *Sophisticated Investors under the Federal Securities Laws*, 1988 DUKE L.J. 1081, 1131 (1988) ("Over time, courts have refined the *Howey* test into three elements: (1) an investment of money (2) in a common enterprise with (3) an expectation of profits that will be derived from others' efforts.").

63. SEC v. W.J. Howey, 328 U.S. 293, 298 (1946). These opinions provide important background and context, but offer little in the way of content not embodied in the *Howey*

the enterprise are evidenced by formal certificates or by nominal interests in the physical assets employed in the enterprise."[64] Moreover, as the *Howey* case illustrates, a unitary legal contract is not required.[65] None of these offers definitive guidance on the meaning of *Howey's* first prong.

The words "transaction" and "scheme" do not have an accepted legal meaning. In common American English usage, the word "transaction" means "something transacted; especially: an exchange or transfer of goods, services, or funds;"[66] the word "scheme," in this context, means "a plan or program of action; especially: a crafty or secret one."[67] These terms have expansive meanings, and the Court tacitly embraced them in the *Howey* opinion in the description of its overall "investment contract" definition:

> It permits the fulfillment of the statutory purpose of compelling full and fair disclosure relative to the issuance of "the many types of instruments that in our commercial world fall within the ordinary concept of a security." It embodies a flexible rather than a static principle, one that is capable of adaptation to meet the countless and variable schemes devised by those who seek the use of the money of others on the promise of profits.[68]

While some crowdfunding business models may not involve the execution of a formal written contract, the sale of an interest in the funded venture typically does involve an offer, an acceptance, mutual obligation, the transfer of consideration and other elements of a legally valid, binding, and enforceable contract under U.S. law.[69] For example, 33needs describes

---

opinion.

64. *Id.* at 299.

65. *Id.* at 300 (noting that the plaintiffs' "respective shares in this enterprise are evidenced by land sales contracts and warranty deeds."); *see also* SEC v. Edwards, 540 U.S. 389, 391 (2004) (involving the sale of payphones "packaged with a site lease, a 5-year leaseback and management agreement, and a buyback agreement"); Hocking v. Dubois, 885 F.2d 1449 (9th Cir. 1989) (involving the purchase of a condominium packaged with several rental agreements).

66. MERRIAM-WEBSTER'S COLLEGIATE DICTIONARY 1327 (11th ed. 2003). Similarly, *Black's Law Dictionary* defines a "transaction" as: "(1) The act or an instance of conducting business or other dealings; . . . . (2) Something performed or carried out; a business agreement or exchange. (3) Any activity involving two or more persons." BLACK'S LAW DICTIONARY 1635 (9th ed. 2009).

67. MERRIAM-WEBSTER'S COLLEGIATE DICTIONARY 1110 (11th ed. 2003). *Black's Law Dictionary* defines a "scheme" as: "(1) A systematic plan; a connected or orderly arrangement, esp. of related concepts." BLACK'S LAW DICTIONARY, *supra* note 66, at 1462.

68. *Howey*, 328 U.S. at 299 (citation omitted).

69. The five primary requirements of a valid contract are:

> one or more promisors and one or more promisees having legal capacity to enter into the contract; at least one promisor or one promisee who is an individual party on but one side of the contract, as distinguished from membership in an entity

the investment[70] process in the frequently asked questions (FAQ) part of its website:

> **How do I invest?**
>
> So simple: just click the big invest button on any company page. You'll be asked to select your investment amount. From there, you will go through Amazon.com's secure and uber simple checkout process.[71]

Accordingly, the investment is made in the form of a standard e-commerce purchase and sale accomplished by either paying funds from a checking account or using a debit or credit card. Under U.S. law, Internet purchases generally are acknowledged to be valid, binding, and enforceable contracts if made by persons having legal capacity.[72] Interestingly, the 33needs website expressly raises questions about both extraterritoriality and the legal capacity of investors:[73]

> **Can I invest if I'm not in the US?**
>
> Absolutely. You can invest (or list your company) if you live on a remote island off the coast of Kenya or live in the heart of New York City.
>
> . . .

> which may be a party on the other side; a manifestation of mutual assent by the parties who form the contract, to the terms of the contract and by each promisor to the consideration for a promise, with limited exceptions; sufficient consideration, again with limited exceptions; and a requirement that the transaction must not be one declared void by statute or by special rules of the common law.

RICHARD A. LORD, 1 WILLISTON ON CONTRACTS § 3:2, at 270–72 (4th ed. 2007) (footnotes omitted); *see also* 17A AM. JUR. 2D *Contracts* § 19 (2004) ("The elements of a valid contract have been stated as . . . an offer, acceptance, contractual capacity, consideration, a manifestation of mutual assent, and legality of the object and of the consideration.").

70. The use of the word "investment" is not meant to convey a legal conclusion for purposes of the *Howey* test or otherwise. Rather, it is the term used for purchases of interests on the 33needs website. *See FAQ*, 33NEEDS, http://www.33needs.com/pages/faq (last visited Sept. 2, 2011) (noting in particular the response to "I'm confused. Why do you call this an investment?").

71. *FAQ*, 33NEEDS, *supra* note 70.

72. *See* JEFFREY H. MATSUURA, SECURITY, RIGHTS, AND LIABILITIES IN E-COMMERCE 185 (Artech House 2001). *See generally* U.C.C. § 2-204(4)(b) (2004) ("A contract may be formed by the interaction of an electronic agent and an individual acting on the individual's own behalf or for another person.").

73. Again, "investors" is the word used on the 33needs website and its use is not intended to convey a legal conclusion. *See FAQ*, 33NEEDS, *supra* note 70.

Do investors need to be a certain age?

Investors who participate can be any age.

Do investors need to be in the U.S.?

Not at all.[74]

Regardless of whether the purchase of crowdfunding interests constitutes a valid, binding, and enforceable contract, the purchase by funders of a crowdfunding interest qualifies as a transaction or scheme—a plan or program to finance a business or project through the exchange of funds for profit-sharing interests. Even before *Howey*, the Court noted that the inclusion of terms such as "investment contract" in the Securities Act evidences a legislative intent to bring "novel, uncommon, or irregular devices" under the coverage of the Securities Act in some circumstances.[75] While crowdfunding is a fairly distinctive and new phenomenon, it is difficult to fathom how a financing plan or program that involves the exchange of funds for profit-sharing interests in a third-party's venture over the Internet would not qualify as a contract, transaction, or scheme under the *Howey* test.

2. Investment of Money

A number of crowdfunding business models involve people spending money with the prospect of getting more money back, seemingly satisfying the second prong of the *Howey* test. As shown in Table 1, many crowdfunding websites offer funders a financial return on the interests purchased.[76] When such a return is offered, it is usually in the form of a

---

74. *Id.*

75. SEC v. C.M. Joiner Leasing Corp., 320 U.S. 344, 351 (1943); *see also* SEC v. W.J. Howey Co., 328 U.S. 293, 298 (1946) ("By including an investment contract within the scope of [Section] 2(1) of the Securities Act, Congress was using a term the meaning of which had been crystallized by . . . prior judicial interpretation," under which state courts had construed the term broadly "so as to afford the investing public a full measure of protection.").

76. There are, however, many crowdfunding websites that do not offer traditional financial returns (in the form of capital appreciation or a participation in earnings) to funders. For example, Kiva is a microfinancing venture that secures loans for start-up businesses in lesser-developed countries. *See How Kiva Works*, KIVA, http://www.kiva.org/about/how (last visited Dec. 26, 2011). The funders receive repayments of the principal of their loans with no interest. *Id.* Additionally, VenCorps provides start-up capital and enables funders to earn points for their contributions, which can be redeemed for various non-financial goods and services. *FAQ*, VENCORPS, http://www.vencorps.com/Page/FAQ#funder (last visited Dec. 26, 2011). For example, the VenCorps store might offer an iPad to be auctioned off to funders who bid points for the product. *Id.* This type of funding model is

revenue-sharing or profit-sharing arrangement.[77] Examining this bargain in isolation, there appears to be a clear income-seeking intent and motive on the part of funders.

However, dicta in the *Howey* case, which was subsequently cited and applied by the Court in *Forman*, raises a question as to whether all exchanges of money made in the hopes of getting a return are investments of money for the purpose of determining the existence of an investment contract.[78] Specifically, the *Howey* Court noted that the purchasers of the land and contracts were "attracted solely by the prospects of a return on their investment."[79] Many crowdfunding websites raise funds to support the production of goods and services by artists and others, and these crowdfunded ventures often reward funders with free or discounted products or services created or sold by the funded business.[80] For example, in May 2011, 33needs featured a business venture, More than Me, Inc., that "funds the education of girls in Liberia by selling laptop covers, which are made by local Liberian women."[81] On the 33needs website, More than Me promises investors a return of "5% of our revenue for 1 year. . . . If you invest $100 or more, you'll get one of the first 100 laptop slip covers."[82] 33needs also plans to offer other benefits to investors based on the number of "impact points" investors earn in funding featured businesses.[83] The

---

less likely to be an investment contract under *Howey*. We question, however, whether the regulation of crowdfunding should turn on this difference.

77. We believe that it is inconsequential for purposes of the *Howey* test whether the current return on a crowdfunding interest represents a share of revenues or profits. For example, one way in which the *Forman* Court referred to profits was as "a participation in earnings . . . ." United Hous. Found., Inc. v. Forman, 421 U.S. 837, 852 (1975). Additionally, the *Edwards* Court found that a fixed rate of return could support investment contract status under *Howey*. *See* SEC v. Edwards, 540 U.S. 389, 394 (2004) ("There is no reason to distinguish between promises of fixed returns and promises of variable returns for purposes of the test, so understood. In both cases, the investing public is attracted by representations of investment income . . . .").

78. *See Forman*, 421 U.S. at 852; *Howey*, 328 U.S. at 300.

79. *Howey*, 328 U.S. at 300.

80. For example, SellaBand is a crowdfunding site that supports music artists and entitles funders to "[r]eceive free downloads and other goodies artists might offer like exclusive CDs, t-shirts, free lunches etc." *How It Works*, SELLABAND, https://www.sellaband.com/en/pages/how_it_works (last visited Dec. 26, 2011). Peerbackers is another website that funds start-up businesses where funders receive various "rewards or perks" at the entrepreneurs' discretion. *FAQ*, PEERBACKERS, http://peerbackers.com/faq (last visited Dec. 26, 2011). Because financial rewards are strictly prohibited, the rewards are usually in the form of products created by the enterprise. *Id.*

81. *More Than Me*, 33NEEDS, http://www.33needs.com/ventures/more-than-me (last visited Sept. 2, 2011).

82. *Id.*

83. *FAQ*, 33NEEDS, *supra* note 70.

hybrid nature of the benefits offered on some crowdfunding websites blurs the line between investment and consumption as set forth in *Forman*.[84]

The 33needs venture raises additional concerns about the investment of money under *Howey* and *Forman*. 33needs focuses on social enterprise funding and markets itself as a financing venture for social entrepreneurs:[85]

> **Just so I'm clear, who raises money on 33needs?**
>
> Social entrepreneurs. They lead companies solving the world's biggest needs, also known as social enterprises. You've heard of them, right? If not, just Google it.[86]

As demonstrated above, 33needs encourages funders to finance the ventures it features based on more than a classic investment or consumption interest. It markets and sells altruism (funding a better world) and meaningfulness (the emotional satisfaction of having an individual impact):

> **This is just about the money, then?**
>
> Not at all. We also believe 33needs powers something as—or even more—important than money: it powers community. And particularly in the world of social good, where people feel an emotional attachment to your company, building a community of passionate supporters (that means you) can often drive change. And that, after all, is what it's all about.
>
> **What's the larger point, though?**
>
> Put simply: it's all about maximizing social and environmental impact. Nothing else matters. Thousands of entrepreneurs need a disruptive financial innovation: a way to connect with the vast pool of capital in the hands of our friends, our family, and the growing number of people who believe business-led solutions allow for a deeper, more sustained impact than old models. We think impact investing can play a crucial role in solving our world's biggest needs. That's why we're here.[87]

This mixed-motivation solicitation likely encourages people to fund businesses featured on the 33needs website for one or more reasons—financial return, preferential access to goods or services, emotional satisfaction, or contribution to the public good.

The hybrid nature of the motivation of purchasers of crowdfunding interests is, however, unlikely to change the conclusion that crowdfunding interests represent an investment of money. In deciding subsequent

84. *See supra* note 32 and accompanying text.

85. *See, e.g.*, *Mission*, 33NEEDS, http://www.33needs.com/pages/mission (last visited Sept. 2, 2011).

86. *FAQ*, 33NEEDS, *supra* note 70.

87. *Id.*

investment contract cases, federal courts generally have given little effect to the language in *Howey* referencing investors attracted "solely by the prospects of a return on their investment."[88] For example, in *Teague v. Bakker*, a case involving "approximately 160,000 individuals who purchased 'Lifetime Partnerships' from an entity known as 'PTL' entitling them to a short stay annually in a hotel at a vacation retreat constructed by PTL,"[89] the court did not give dispositive weight to testimony from some purchasers that their motivation was personal use of the hotel rather than profit:

> It would make little sense for the existence of a "security" to turn solely on whether those who actually invest do so without regard to profit. Such a rule would be highly impractical. Would the existence of a "security" change according to each purchaser? If not, how many, or what percentage of, purchasers would have to have made their investments with an eye toward profits in order for there to be "securities"? Finally, how could the SEC be expected to regulate effectively where the existence of a "security" turns not on how and to whom an investment opportunity is offered, but only on those who ultimately undertake such an investment?[90]

In *Teague*, the court relied on the promotional materials used to solicit purchases to find that the Lifetime Partnerships may represent an investment of money with the expectation of profit.[91] Interestingly, the *Howey* Court offered a similar analysis in determining that an investment contract existed:

> This conclusion is unaffected by the fact that some purchasers choose not to accept the full offer of an investment contract by declining to enter into a service contract with the respondents. The Securities Act prohibits the offer as well as the sale of unregistered, non-exempt securities. Hence it is enough that the respondents merely offer the essential ingredients of an investment contract.[92]

Crowdfunding websites that offer returns to their funders promote the revenue-sharing or profit-sharing components of that return. For example, each of the featured businesses on 33needs has its own page on the 33needs website that includes an "offer" to the investors.[93] This page is where the return on investment is described. In addition, the FAQ page on the 33needs website clearly distinguishes the capital investments it seeks from donations

88. SEC v. W. J. Howey Co., 328 U.S. 293, 300 (1946) (emphasis added).
89. Teague v. Bakker, 35 F.3d 978, 981 (4th Cir. 1994).
90. *Id.* at 988 n.12.
91. *Id.* at 988–89.
92. *Howey*, 328 U.S. at 300–01 (footnote omitted).
93. Venture Listing, 33NEEDS, http://www.33needs.com/ventures.

and loans, noting that financial rewards are a fundamental, unique part of the bargain in funding one of its ventures:[94]

> We're turning the focus of crowdfunding entirely on companies with a social mission, and allowing ordinary people to invest, make a social impact, and *earn financial rewards.*
>
> *We're the only platform in the world that enables ordinary people – you, neighbor Joe, Aunt Sally, and even your dog Jake (fine, maybe not Jake) – to invest in do-good companies (called social enterprises) and earn financial rewards.* For example: 20,000 people could invest, not donate, $50 each in More than Me, Inc., one of the companies raising money on 33needs.[95]

This type of marketing supports a conclusion that the purchase of the subject crowdfunding interests constitutes an investment of money.

Some cases analyzing investment contracts have focused on whether the primary, rather than exclusive, purpose of the arrangement is to provide a return to funders.[96] Accordingly, in the case of crowdfunding websites or crowdfunded ventures that offer both non-financial benefits and financial return, the satisfaction of the *Howey* test may depend upon whether the primary purpose of the arrangement is affording funders preferential access to goods or services, offering them emotional satisfaction, presenting them with an opportunity to contribute to the public good, or providing them a financial return. Although there are variations among the financing arrangements on crowdfunding websites we have reviewed, the potential financial return on the crowdfunding interests described in Table 1 as offering "hybrid returns" may have a significantly higher potential value than the non-financial interest offered.[97] This value disparity may

94. *FAQ*, 33NEEDS, *supra* note 70.

95. *Id.* (emphasis added).

96. *See* Int'l Bhd. of Teamsters v. Daniel, 439 U.S. 551, 560 (1979) (holding that involvement in a mandatory pension plan was not an investment contract). The Court stated, "[l]ooking at the economic realities, it seems clear that an employee is selling his labor *primarily* to obtain a livelihood, not making an investment." *Id.* (emphasis added); *see also* Aschenbach v. Covenant Living Centers-North, Inc., 482 F. Supp. 1241, 1244 (E.D. Wis. 1980) ("[R]esidency contracts, which are not transferrable or assignable, are entered into for the primary purpose of acquiring low cost living space, with the added feature of low cost maintenance and health care, and not for 'profit' in the sense intended by the Supreme Court in *Howey* and *Forman*.").

97. For example, SellaBand offers a revenue-sharing arrangement (at the artist's discretion) as well as music downloads, CDs, and t-shirts signed by or promoting the artist. *How it Works*, SELLABAND, *supra* note 80. However, there is no limit on the amount that an individual funder may contribute. This creates the potential for huge differences between the monetary values of the financial and non-financial interests funders may receive. For example, as a result of a large contribution, a funder on SellaBand could be entitled to a large percentage of revenue, which would have a much greater value than the non-financial

encourage funders to purchase crowdfunding interests, rather than, for example, buy goods or services from a traditional brick-and-mortar or online retailer, where a consumption interest also is involved or contribute to a charity that funds a social project, where altruism or the public good is an objective. In other words, a higher potential value of the financial rewards promised on a crowdfunding website may support the conclusion that the site is primarily offering a revenue-sharing or profit-sharing arrangement. Under these circumstances, it is hard to argue that the funder's primary purpose in purchasing crowdfunding interests on these websites is not the investment of money.

Finally, in determining whether a contract, transaction, or scheme represents an investment of money, some courts focus on whether the arrangement subjects the funder to a loss.[98] Under this analysis, which is sometimes seen as the equivalent of a risk capital analysis,[99] crowdfunding interests may represent a form of financing that subjects the funder to the loss of his or her initial investment. Crowdfunding interest purchasers typically have little or no control over the success of the business they fund, and unless funds are conveyed in the form of a debt instrument, have no right to a return of their capital (although many equity-type crowdfunding

---

interests (i.e., music downloads, CD, t-shirts, etc.).

98. *See, e.g.*, Becks v. Emery-Richardson, Inc., No. 86-6866, 1990 U.S. Dist. LEXIS 21066, at *36 (S.D. Fla. July 6, 1990) ("Generally, an 'investment' in this context means that the investor commits his assets to an enterprise or venture in such a manner as to subject himself to financial loss.").

99. *See, e.g.*, Underhill v. Royal, 769 F.2d 1426, 1431 (9th Cir. 1985) (outlining a Ninth Circuit risk capital test that was used to determine whether notes are securities prior to the Court's opinion in *Reves v. Ernst & Young*, 494 U.S. 56 (1990)). In the investment contract context under our framework, the risk capital test is typically used to assess the combination of the second, fourth, and fifth prongs of the *Howey* test. First Citizens Fed. Sav. & Loan Ass'n v. Worthen Bank & Trust Co., N.A., 919 F.2d 510, 516 (9th Cir. 1990). Given its substantial overlap with key parts of the *Howey* test, the risk capital test sometimes is seen as an alternative to the *Howey* test. *See, e.g.*, Martin v. T. V. Tempo, Inc., 628 F.2d 887, 891 (5th Cir. 1980) ("Plaintiffs urge that the district court erred in failing to analyze the franchise agreement under the so-called 'risk capital' approach. We previously have taken note of this alternative to the *Howey* test . . . ."). In its modern formulation, the "risk capital" test examines four factors:

> The "risk capital" test requires a consideration of the following factors: (1) whether funds are being raised for a business venture or enterprise; (2) whether the transaction is offered indiscriminately to the public at large; (3) whether the investors are substantially powerless to effect the success of the enterprise; and (4) whether the investors' money is substantially at risk because it is inadequately secured.

Moreland v. Dep't of Corp., 239 Cal. Rptr. 558, 566 (Cal. App. 5th Dist. 1987) (decided under California State law).

business models provide that capital will be returned if a stated funding threshold is not met[100]).

Based on the promotion of a money-making potential, crowdfunding interests that offer financial returns to funders likely satisfy the second prong of our five-pronged *Howey* test.

3. Common Enterprise

Turning to the "common enterprise" part of the *Howey* test, we begin with horizontal commonality, noting that the pooling of funds obtained from the crowd is seemingly the essence of crowdfunding.[101] Funders purchase interests in a particular crowdfunded venture featured on a crowdfunding website, and typically earn financial returns through revenue-sharing or profit-sharing, based on the amount of their investment as a percentage of the business's aggregate funding target. Their ability to benefit financially arises from the success of the overall venture. The investor's funds are locked in once the venture reaches its funding target. 33needs calls this aspect of its operations "all or nothing funding":

> **What if the company doesn't hit its funding target?**
>
> It's all or nothing funding. For example, if a company attempts to raise $20,000 and falls short of their target, then all the investors will get their money back.
>
> **Is that "all for nothing" model fair to these good companies?**
>
> We just think it's less risky for everyone. We want you, the investor, to have confidence that the company has raised enough to completely follow-through on their commitments.[102]

This basic financing arrangement, common to many of the crowdfunding websites we have reviewed, satisfies the requirements of horizontal commonality.

With respect to vertical commonality, only broad vertical commonality exists in the typical crowdfunding model. Investor success is generally dependent on the combined efforts of the crowdfunding website and the crowdfunded venture.[103] In most cases, a crowdfunding website (one that is unaffiliated with the crowdfunded ventures it promotes) takes a fee or

---

100. *See infra* note 102 and accompanying text.

101. *See supra* notes 2–3 and accompanying text. The 33needs website references the pooling of funds, classifying crowdfunding as "the collective cooperation by people who network and pool their money together." *Mission*, 33NEEDS, *supra*, note 85.

102. *FAQ*, 33NEEDS, *supra* note 70.

103. *See supra* text accompanying note 44.

commission once the featured venture's funding target is achieved and has no ongoing interest in the success of that venture parallel to that of the funders.[104] However, if a crowdfunding website takes a percentage of the featured business's revenues or profits along the same lines as the funders, strict vertical commonality may exist.

The 33needs website is silent as to the nature of the compensation or financial benefit, if any, that it receives for promoting the featured ventures, but the principal of 33needs confirmed that 33needs takes a 5% fee from any venture that successfully reaches its funding target.[105] The FAQ portion of the website is clear, however, about the fact that 33needs does not take an ownership or intellectual property interest in the enterprises it features.[106] Accordingly, 33needs exhibits broad vertical commonality only.

Although a case-by-case analysis of individual crowdfunding sites would yield different results with respect to the existence of broad or strict vertical commonality, the fact that almost every crowdfunding site that we examined satisfies the horizontal commonality test, which is sufficient to establish a common enterprise in many jurisdictions,[107] leads to the conclusion that most crowdfunding business models constitute common enterprises for purposes of the *Howey* test.

4. Expectation of Profits

In analyzing the second part of the *Howey* test under our taxonomy—the investment of money prong—we necessarily engaged some of the analysis relevant to this fourth attribute of an investment contract under *Howey*, the expectation of profits. An investor of money is one who is motivated by financial return in making an expenditure of funds.[108] As noted in our description and analysis in Part II.B.2, the financial benefit that funders expect (and are led to expect) from ventures promoted through crowdfunding websites is a participation in the venture's revenues or profits.[109]

---

104. For example, IndieGoGo provides funding for various entrepreneurial projects and causes for which it takes 4% of the funded amount, but retains no on-going interest in the funded venture. *FAQs*, INDIEGOGO, www.indiegogo.com/about/faqs (last visited Dec. 26, 2011). Similarly, Kickstarter provides funding for artists' projects and charges 5% of the funded amount, but the project's success is entirely in the artists' hands. *FAQs*, KICKSTARTER, www.kickstarter.com/help/faq (last visited Dec. 26, 2011).

105. Email message from Josh Tetrick to Joan Heminway, May 23, 2011 2:08 PM (on file with the Tennessee Law Review).

106. *FAQ*, 33NEEDS, *supra* note 70 ("Does 33needs or investors on 33needs take some percentage of ownership or intellectual property in the social enterprises? Absolutely not.").

107. *See supra* notes 35–36 and accompanying text.

108. *See supra* text accompanying notes 77–79.

109. *See supra* Part II.B.2.

There are some crowdfunding websites (typically sites offering debt-type interests) that offer a fixed return to funders.[110] Offering this type of benefit arrangement (in lieu of revenue-sharing or profit-sharing) is sufficient to satisfy the expectation of profits prong of the *Howey* test.[111]

Crowdfunding websites do not typically allow funders to benefit through capital appreciation, the other type of financial return recognized as profit under the *Forman* case.[112] Although most sites do not address the issue, there is no apparent web-based mechanism for transferring crowdfunding interests to others. The interests, however, constitute personal property and therefore should be assignable. Moreover, with minor exception, there is no evidence that a market exists for the transfer or assignment of all or some rights in crowdfunding interests.[113] However, if a market were to develop, the expectation of profits from capital appreciation also would be possible.

Those who purchase crowdfunding interests that promise a current return or capital appreciation expect profits under *Howey*.

5. Solely from the Efforts of Others

The "solely from the efforts of others" prong of the *Howey* test[114] also is met with respect to almost every crowdfunding business. Under many crowdfunding business models, the funder serves as a passive patron while the principals of the crowdfunded venture, with some marketing or logistical support from the crowdfunding website, are responsible for the venture's success or failure.[115] In a few instances we have observed, the funder plays a minor role in selecting, promoting, or conducting the crowdfunded venture.[116] Regardless of the funders' exact level of

110. Microplace is a microfinancing operation that allows funders to contribute money in the form of a loan to entrepreneurs in less-developed countries. *How It Works: Overview*, MICROPLACE, https://www.microplace.com/howitworks (last visited Dec. 26, 2011). The funder is entitled to a return of the funded amount plus interest. *Id.* Similarly, 40Billion provides start-up funds for businesses, on an invitation-only basis, and the funder is entitled to repayment of the principal plus interest. *How It Works*, 40BILLION, www.40billion.com/how_it_works.asp (last visited Dec. 26, 2011).

111. *See supra* note 47 and accompanying text.

112. *See supra* note 46 and accompanying text.

113. One site that aims to facilitate a market for crowdfunding shares is Cinema Shares, which plans to allow for the purchase of fully listed, publicly tradable on NASDAQ. *About Cinema Shares*, CINEMA SHARES, www.cinemashares.com/aboutCS.html (last visited Dec. 26, 2011).

114. SEC v. W. J. Howey Co., 328 U.S. 293, 300 (1946).

115. *See* Belleflamme et al., *supra* note 3, at 3 ("The major fraction are passive investments; i.e., investments with a promise of compensation but no direct involvement in the decision-making process or provision of time or expertise for the initiative.").

116. For example, funders who choose to participate as "believers" on SellaBand also can support the artists of their choice in other ways:

involvement in a crowdfunded venture, the bulk of the efforts contributing to the venture's success is supplied by the principals of the crowdfunded venture, with some support (typically administrative and ministerial) from the crowdfunding website.[117]

The crowdfunding website's level of engagement in the operations of the crowdfunded venture varies in different crowdfunding models. 33needs, for example, pre-screens the businesses that are featured on the site:

> To date, 33needs has received over 900 applications from entrepreneurs around the world. We have a 4-person investment committee that reviews, interviews, and selects companies. Sometimes we select them individually, sometimes collectively. We look at the strength of the business model, integrity of the team, and the nature of the need being addressed. We tend to focus on urgent needs.[118]

In addition, while 33needs states that the funded ventures are responsible for upholding the commitments ("offers") they make to investors,[119] it also promises that it will "be working closely with them to make sure all is right."[120]

In any case, funder profits on the crowdfunding websites we reviewed would result exclusively or primarily from the efforts of the promoters (including the crowdfunding websites and crowdfunded ventures), not from the efforts of the funders themselves. Thus, the fifth and last prong of the *Howey* test, like the other four, likely is satisfied by most crowdfunded ventures. As a result, it is probable that a court would find that crowdfunding interests that include a financial return are investment contracts.

C. The Question of Context

Under Section 2(a)(1) of the Securities Act, an investment contract is a security "unless the context otherwise requires."[121] Section 2's introductory

---

> On SellaBand you can support your favorite artists by buying a part and helping them to raise the funds for a new music project (a new album, tour or the promotion of their music). . . . Join your favorite artists on their way to reach their funding goal. Promote them, stay in touch with them and help them.

How It Works, SELLABAND.COM, *supra* note 80.

117. *Id.*

118. *FAQ*, 33NEEDS, *supra* note 70.

119. *Id.* (noting that the ventures, themselves, are "responsible for making sure [they] deliver on what they promise").

120. *FAQ*, 33NEEDS, *supra* note 70.

121. 15 U.S.C. § 77b(a)(1) (2006).

limitation regarding context is seldom used to disqualify an instrument listed in Section 2(a)(1) from being a security. However:

> courts have held that the definition of what constitutes a security need not be read literally, thereby giving meaning to the introductory language of the definition, "unless the context otherwise requires." Stressing that the Acts were aimed at preventing fraud in the securities market, the Supreme Court has stated that "[b]ecause securities transactions are economic in character Congress intended the application of these statutes to turn on the economic realities underlying a transaction, and not on the name appended thereto." This is the examination which must be made "in searching for the meaning and scope of the word 'security'"—"form should be disregarded for substance and the emphasis should be on economic reality."[122]

As the Supreme Court has explained, "[t]he test . . . is what character the instrument is given in commerce by the terms of the offer, the plan of distribution, and the economic inducements held out to the prospect. . . . [I]t is not inappropriate that promoters' offerings be judged as being what they were represented to be."[123] For example, court opinions assessing the status of "notes" under Section 2(a)(1) have used the context limitation to find that certain notes are not securities for purposes of the Securities Act.[124] The determinative factor in these cases is whether the note is a commercial instrument or an investment instrument.[125] In the same vein, the Supreme Court elevated substance over terminology in *United Housing Foundation v. Forman*[126] when it determined that interests in a housing cooperative were not securities despite the fact that the interests were labeled as "stock," one of the instruments listed in the security definition.[127]

---

122. Ayala v. Jamaica Sav. Bank, No. CV-80-1802, 1981 U.S. Dist. LEXIS 17994, at *9 (E.D.N.Y. June 15, 1981) (citations omitted).

123. SEC v. C. M. Joiner Leasing Corp., 320 U.S. 344, 352–53 (1943).

124. *See* Hunssinger v. Rockford Bus. Credits, Inc., 745 F.2d 484, 487 (7th Cir. 1984) ("[T]his as well as other Circuits have relied upon the prefatory phrase 'unless the context otherwise requires' to exclude certain notes from the protection of the federal securities acts."); *see also* Lincoln Nat'l Bank v. Herber, 604 F.2d 1038, 1043 (7th Cir. 1979) ("Congress itself has cautioned that the same words may take on a different coloration in different sections of the securities laws; both the 1933 and 1934 Acts preface their lists of general definitions with the phrase 'unless the context otherwise requires'").

125. *Hunssinger*, 745 F.2d at 488.

126. United Hous. Found., Inc. v. Forman, 421 U.S. 837 (1975).

127. *Id.* at 851 ("noting that the interests at issue lack what the Court in Tcherepnin deemed the most common feature of stock: the right to receive 'dividends contingent upon an apportionment of profits'" and also fail to "possess the other characteristics traditionally associated with stock: they are not negotiable; they cannot be pledged or hypothecated; they confer no voting rights in proportion to the numbers of shares owned; and they cannot appreciate in value." (citation omitted)).

The key to analyzing context hinges on whether the financial instrument at issue represents a financial investment vehicle. In this regard, the Supreme Court in *Reves* stated, "A commitment to an examination of the economic realities of a transaction does not necessarily entail a case-by-case analysis of every instrument, however. Some instruments are obviously within the class Congress intended to regulate because they are by their nature investments."[128] Because the *Howey* test, as applied, includes an evaluation of a contract, transaction, or scheme as an investment of money, our analysis in Part II.B.2 already establishes the necessary context. Crowdfunding interests that include revenue-sharing or profit-sharing benefits appear to be equity-type capital investment vehicles despite the potential other benefits they may offer (e.g., consumption interests and altruistic and emotional satisfaction).[129] Although we recognize that crowdfunding interests are unique and flexible devices, these attributes do not exempt them from regulation:

> [T]he reach of the Act does not stop with the obvious and commonplace. Novel, uncommon, or irregular devices, whatever they appear to be, are also reached if it be proved as matter of fact that they were widely offered or dealt in under terms or courses of dealing which established their character in commerce as "investment contracts," or as "any interest or instrument commonly known as a 'security.'"[130]

The analysis in this Part II demonstrates that crowdfunding interests that include a financial return are offered under terms and in courses of dealing that establish their character in commerce that are investment contracts and securities under *Howey* and its progeny.

III. FEDERAL REGULATION OF OFFERINGS OF CROWDFUNDING INTERESTS AS SECURITIES

Our analysis in Part II.B indicates that the interests offered to funders by some crowdfunding websites—those that offer a revenue-sharing or profit-sharing arrangement—likely satisfy all five elements of the *Howey* test and, therefore, are investment contracts. Because the context in which crowdfunding interests are offered and sold to funders does not otherwise require a different categorization,[131] we assume for the remainder of this article that these crowdfunding interests are securities within the meaning of Section 2(a)(1) of the Securities Act.[132] The status of these interests as securities exposes crowdfunding websites and crowdfunded ventures to the

128. Reves v. Ernst & Young, 494 U.S. 56, 62 (1990).
129. *See supra* notes 76–96 and accompanying text.
130. SEC v. C. M. Joiner Leasing Corp., 320 U.S. 344, 351 (1943).
131. *See supra* Part II.C.
132. 15 U.S.C. § 77b(a)(1) (2006).

prospect of regulation—including through the expensive and time-consuming process of registering offers and sales of securities—under the Securities Act. This Part describes both the current regulatory framework and the underlying policy objectives of that landscape as applied to crowdfunding interests. With that framework and landscape in mind, Part III concludes by questioning whether crowdfunding interests that are securities should be subject to the registration requirements of the Securities Act.

A. Applicable Regulation

A full-blown description of regulation under the U.S federal securities laws is beyond the scope of this article (and, indeed, is the subject of three-credit-hour-plus courses in Securities Regulation in U.S. law schools). However, even a brief summary of key applicable provisions of the federal securities laws (which is what we provide here) illustrates the weight of regulation they impose—a transaction cost that is impossible for small businesses to bear.

1. Registration and Related Liability and Costs

Section 5 of the Securities Act regulates the offer and sale of securities.[133] In sum, Section 5 prohibits the offer or sale of securities without registration, unless an applicable exemption is available.[134] An "offer for sale" or "sale" includes "every attempt or offer to dispose of, or solicitation of an offer to buy, a security . . . ."[135] These terms are interpreted very broadly, making it quite easy for an offeror or seller of securities to inadvertently violate Section 5 by, for example, communicating with potential investors before filing a registration statement.[136] If interests in crowdfunded ventures are securities, then the offer and sale of those interests through a crowdfunding website must be registered with the SEC, absent an applicable exemption. A parallel system of registration exists under state law.

If securities are offered or sold in violation of Section 5, Section 12(a)(1) provides the securities purchaser with a private cause of action

133. 15 U.S.C. § 77e.

134. *Id.*

135. 15 U.S.C. § 77b(a)(3).

136. *See* Stephen J. Choi, *Company Registration: Toward a Status-Based Antifraud Regime*, 64 U. CHI. L. REV. 567, 606 (1997) ("Section 5 sweeps broadly, regulating every offer and sale of a security."); Joseph F. Morrissey, *Rhetoric and Reality: Investor Protection and the Securities Regulation Reform of 2005*, 56 CATH. U. L. REV. 561, 568 (2007) ("Section 5(c) of the Securities Act specifically made it unlawful for any person to offer to sell or buy securities before a registration statement had been filed with the SEC." "Section 2(a)(3) of the Securities Act defined "offer" as broadly as it could . . . .").

against the seller, allowing for rescission of the sale, or recovery of rescissory damages if the purchaser no longer owns the securities.[137] False and misleading registration statements are actionable under Section 11 of the Securities Act,[138] false and misleading prospectuses or oral communications may result in liability under Section 12(a)(2) of the Securities Act,[139] and fraudulent conduct in connection with the offer and sale of securities may be enforced (at least by the SEC) under Section 17(a) of the Securities Act.[140] Further, those who offer and sell securities are exposed to potential liability for securities fraud claims under Section 10(b) of the Securities Exchange Act of 1934, as amended (the "Exchange Act"),[141] and Rule 10b-5 under the Exchange Act.[142]

This is a heavy system of regulation. Registration of the offer and sale of securities under the Securities Act is an expensive and time-consuming proposition, and the prospect of lengthy, costly enforcement actions by private plaintiffs (in individual or class actions), the SEC, and the Department of Justice loom large. As for the federal registration requirements, an issuer must file a registration statement that includes operating and financial disclosures about the issuer, information about the securities being offered and sold, and details about the plan of distribution of those securities.[143] The costs of an initial SEC registration typically include underwriting compensation, a registration fee paid to the SEC, legal and accounting fees and expenses, printing and engraving costs, a Financial Industry Regulatory Authority filing fee, electronic filing fees when using a service for filing, stock exchange listing fees (if applicable), Blue Sky filing fees (if applicable), and transfer agent and registrar fees when the issuer retains the services of a third party to handle its stock records.[144] Although

137. 15 U.S.C. § 77l(a)(1).

138. 15 U.S.C. § 77k.

139. 15 U.S.C. § 77l(a)(2).

140. 15 U.S.C. § 77q(a); *In re* Washington Pub. Power Supply Sys. Sec. Litig., 823 F.2d 1349, 1353–54 (9th Cir. 1987).

141. 15 U.S.C. § 78j(b). Although the Exchange Act has its own definition of the term "security," the definitions under the Securities Act and the Exchange Act are substantially similar, and result in only small differences in application. *See* JAMES D. COX ET AL., SECURITIES REGULATION: CASES AND MATERIALS 19 (6th ed. 2009) ("The '33 Act and the '34 Act have substantially similar definitions of a *security*."). Furthermore, the regulatory schemes of the two acts are integrated, with the Securities Act regulating offers and sales of securities generally and the Exchange Act largely governing trading transactions on and through securities markets and market professionals. *See* COX ET AL., *supra*, at 7 ("Whereas the Securities Act grapples with the protection of investors in primary distributions of securities, the Exchange Act's concern is trading markets and their participants.").

142. 17 C.F.R. § 240.10b-5 (2011).

143. *See* 15 U.S.C. § 77g; COX ET AL., *supra* note 141, at 143–47.

144. *See* COX ET AL., *supra* note 141, at 156 (setting forth in Note 4 various external and internal costs of going public); Stuart R. Cohn & Gregory C. Yadley, *Capital Offense: The SEC's Continuing Failure to Address Small Business Financing Concerns*, 4 N.Y.U. J.

certain offerings cost somewhat less because of the nature of the issuer or the offering (and offering costs are significantly higher for large and complex offerings), an initial public offering for even a small business will cost the issuer over $100,000 in fees for third-party services alone, and this figure does not include the value of the time senior management spends preparing for and marketing the offering.[145] Small business issuers may have lower registration costs than large businesses in some regards, but the overall relative costs are high:

> [O]ne should appreciate that it is relative, not absolute, offering expenses that are important. To use an extreme example, $500,000 in offering costs on a $50 million offering will certainly not kill the transaction, while $500,000 in offering expenses on a $500,000 deal will kill the transaction.
>
> Accounting, legal and other expenses on small deals can easily exceed $50,000, and such amounts bulk large relative to the total yield from a small offering. When added to the costs due to the lack of financial intermediation services, one is able to appreciate the extreme structural and economic disadvantages that small entrepreneurs encounter when attempting to access external capital.[146]

Add to these expenses the ongoing costs of being public, which, depending on the issuer's assets and equity ownership, may be a long-term

---

L. & BUS. 1, 8–9 (2007) ("Moreover, regardless of the outcome of the offering, the costs of the registration process are heavily front-loaded. Accounting fees, attorney retainers, SEC filing fees, broker-dealer expenses, printing and road show costs are all incurred and become payable prior to the effective date of the registration statement.").

145. One commentator accurately sums up the cost situation:

> Registration involves legal fees, accounting fees, printing costs, filing fees, and other miscellaneous costs, along with a significant discount paid to the underwriters. The total expense is hundreds of thousands of dollars. These external costs are in addition to the time consumed by the company's own employees in preparation for registration.

C. Steven Bradford, *Securities Regulation and Small Business: Rule 504 and the Case for an Unconditional Exemption*, 5 J. SMALL & EMERGING BUS. L. 1, 24 (2001); *see also* COX ET AL., *supra* note 141, at 156 ("The estimated 2007 costs for a significant IPO are $600,000–$800,000 in fees to counsel, $400,000–$600,000 for the auditor, underwriter commissions of typically 7 percent of the offering amount, $150,000–$200,000 in printing costs, plus various filing fees . . . ."); Marvin E. Rooks, *It Is Time for the Federal Trade Commission to Require Financial Performance Representations to Prospective Franchisees*, 11 WAKE FOREST J. BUS. & INTELL. PROP. L. 55, 66 (2010) ("The SEC's initial public offering . . . process for even a small company (less than $20 million in revenue) takes six to nine months and costs at least $100,000 in fees for legal, accounting, audit, printing, filing fees, and underwriter commissions." (footnote omitted)).

146. Rutheford B Campbell, Jr., *Regulation A: Small Businesses' Search For "A Moderate Capital,"* 31 DEL. J. CORP. L. 77, 90 (2006).

proposition,[147] and a small business issuer will typically find that the costs of a registered public offering (even without taking into account the prospect of private and public enforcement actions) outweigh the benefits.[148] Registration typically takes several months (at a minimum) because of the length and complexity of the registration statement, the regulatory filing and review process, and the marketing and sales activities.[149] As a result, issuers may miss important financing opportunities because favorable market conditions for an offering (so-called "market windows") will pass unutilized if the offering's registration statement has not yet been declared effective. Missing a market window can be especially devastating to small business issuers who can illafford to lose the sunk costs expendedin initiating and completing the registration process.

The costliness and protracted nature of the registration process are unfortunate because the registration process has a number of advantages for small businesses (as well as other issuers):

> If registration were an economically viable alternative for small issuers, it would produce a number of attractive benefits. It would ameliorate problems of inadvertent loss of an exemption through the impact of the integration doctrine or failure to meet the technical requirements of a particular exemption. It would eliminate all resale restrictions that often adversely impact the attractiveness of exemptions. Finally, it would provide some help and comfort regarding antifraud compliance. Scheduled disclosure requirements in registration forms provide a prepackaged checklist regarding matters and events that may be material and thus subject to disclosure obligation under antifraud rules, such as Rule 10b-5. Compliance with the registration form, therefore, effectively

147. Cohn & Yadley, *supra* note 144, at 9 ("Once public, the company is now subject to the periodic reporting obligations of the . . . Exchange Act . . . for at least the remainder of the first year. . . . These reporting and regulatory burdens weigh extraordinarily heavily on public-traded small businesses, prompting both administrative and legislative efforts to modify such requirements for small business issuers." (footnote omitted)); *see* COX ET AL., *supra* note 141, at 156 ("The publicly traded company incurs the burden of complying with the periodic reporting requirements of the '34 Act. While out-of-pocket costs may be trivial in relation to the registrant's assets or income, the more significant costs are those associated with the consciousness of operating in the public eye.").

148. *See* Campbell, *supra* note 146, at 91–92 ("Registration has never been a viable way for small businesses to raise capital. High transaction costs associated with registered offerings inevitably put registration out of the range of small businesses in search of capital. Thus, the data show that small offerings are very rarely made through SEC registration." (footnotes omitted)); Cohn & Yadley, *supra* note 144, at 10 ("The combined effect of the costs imposed by the registration process and the post-registration reporting system is generally more than sufficient to convince small businesses that financing through a registered public offering is a most undesirable course.").

149. *See* Bradford, *supra* note 4, at 27–28.

reduces the risk of a material omission of fact that would generate liability under federal antifraud rules.[150]

Furthermore, with the SEC's recent approval of the NASDAQ OMX BX proposal to establish a new listings market, the "BX Venture Market,"[151] registration may afford some small business issuers the prospects of accessing a formal public trading market. This market is currently anticipated to launch in 2012.[152] Listed companies will be required to comply with state, as well as federal, securities laws relating to the offer and sale of securities.[153]

2. Exemptions from Registration

Despite its regulatory and potential practical advantages, registration is a nonstarter for most crowdfunding websites and crowdfunded ventures because of the expense and prolonged nature of the process. For crowdfunding websites and crowdfunded ventures, as with many other small businesses, the amount of money and time required to register a securities offering will most often prohibit the offering from occurring.[154] Therefore, under the current regime, the only practical means by which

---

150. Campbell, *supra* note 146, at 92 n.55 (citations omitted); *see also* Stuart R. Cohn, *The Impact Of Securities Laws On Developing Companies: Would the Wright Brothers Have Gotten Off the Ground?*, 3 J. SMALL & EMERGING BUS. L. 315, 361 (1999) ("The SEC finds two principal benefits from registration—disclosure and the ability to resell securities."). *But see* Bradford, *supra* note 145, at 28–29 (describing inconclusive benefits of registration).

151. Self-Regulatory Organizations; NASDAQ OMX BX, Inc.; Order Granting Approval of Proposed Rule Change, Securities Exchanged Act Release No. 34-64437 (May 6, 2011), *available at* http://www.sec.gov/rules/sro/bx/2011/34-64437.pdf.

152. *See FAQ: When will the BX Venture Market launch?*, BX VENTURE MARKET, http://www.bxventure.com/faq (last visited Dec. 26, 2011).

153. *See FAQ: How does the BX Venture Market compare to the NASDAQ Stock Market?*, BX VENTURE MARKET, http://www.bxventure.com/faq (last visited Dec. 26, 2011).

154. *See generally* Cohn & Yadley, *supra* note 144, at 10–15 (discussing how the current regulatory regime fails to adequately provide opportunities for small businesses); Schapiro Testimony, *supra* note 11, at 1 ("Cost-effective access to capital for companies of all sizes plays a critical role in our national economy, and companies seeking access to capital should not be overburdened by unnecessary or superfluous regulations."). We note, however, that at least one crowdfunded business has pursued the registration of a crowdfunded offering. *See* Audience Prod., Inc., Amendment No. 7 to Form S-1 Registration Statement (Form S-1/A) (Apr. 21, 2010), *available at* http://www.sec.gov/Archives/edgar/data/1474227/000147422710000015/ds1a.htm. Ultimately, despite extensions of the originally established offering period, this offering was unsuccessful. In a post-effective amendment filed in August 2011, Audience Productions requested deregistration of the shares offered. *See* Audience Prod., Inc., Post-Effective Amendment No. 6 to Form S-1 Registration Statement (Form S-1/A) (Aug. 8, 2011), *available at* http://www.sec.gov/Archives/edgar/data/1474227/000147422711000018/dposam.htm.

crowdfunding websites and crowdfunded ventures can offer or sell securities is to find an applicable exemption for the security or offering.

Securities may be exempt under Section 3(a) of the Securities Act.[155] For example, Section 3(a) of the Securities Act exempts securities issued by states and municipalities, charitable organizations, and savings and loan associations.[156] Section 3(a) does not currently provide an exemption for crowdfunding interests.

The few possible transactional registration exemptions under the Securities Act that one would consider in connection with a primary offering of interests in a crowdfunded business include: the private offering exemption under Section 4(2);[157] Rules 504, 505, and 506 of Regulation D;[158] and Regulation A.[159] However, none of these exemptions provides a feasible path for a crowdfunding website or crowdfunded venture to avoid registering the offer or sale of profit-sharing interests in the crowdfunded venture.[160] Part III.A.2 outlines the key attributes of each of these possible exemptions and comments on the unsuitability of each for primary offerings of crowdfunding interests.

a. Private Offering Exemption under Section 4(2) of the Securities Act

Section 4(2) of the Securities Act exempts from registration "transactions by an issuer not involving any public offering."[161] Interestingly, the term "public offering" is not defined in the Securities Act or in SEC rules under the statute.[162] However, it is generally acknowledged that the exemption "was designed to apply to specific or isolated sales as well as offerings to a very small number of securities holders so that the public interest is not involved."[163]

---

155. 15 U.S.C. § 77c(a) (2006).

156. *Id.* § 77c(a)(2), (4), (5).

157. *Id.* § 77d(2).

158. 17 C.F.R. §§ 230.501–508 (2011) (authorized under Sections 3(b) and 4(2) of the Securities Act).

159. 17 C.F.R. §§ 230.251–263 (authorized under Section 3(b) of the Securities Act).

160. While the intrastate offering exemption, 15 U.S.C. § 77c(a)(11), may be applicable in some situations involving crowdfunding, most crowdfunded ventures seek to raise capital from investors residing in various states. Because of its unlikely applicability in this context, we do not further analyze the possible application of the intrastate offering exemption in the crowdfunding context.

161. 15 U.S.C. at § 77d(2).

162. *See* Stephen D. Bohrer, *The Application of U.S. Securities Laws to Overseas Business Transactions*, 11 STAN. J.L. BUS. & FIN. 126, 153 (2005) ("The term 'public offering' is not defined under the Securities Act."); Patrick Daugherty, *Rethinking the Ban on General Solicitation*, 38 EMORY L.J. 67, 71 (1989) ("The term 'public offering' is not defined by statute . . . ." (footnote omitted)).

163. THOMAS LEE HAZEN, THE LAW OF SECURITIES REGULATION § 4.24 (5th ed. 2005).

Soon after the Securities Act was signed into law, the SEC's General Counsel set forth five attributes of offerings that provide guidance on whether the offer or sale of securities is a transaction not involving a public offering under Section 4(2): the number of offerees, the relationship of offerees to each other and to the issuer, the number of units offered, the size of the offering, and the manner of the offering.[164] The overall message was that a securities offering is more likely to be characterized as a public offering if:

- the securities are offered to the many (rather than the few),
- the securities are offered to those with no or little preexisting association to each other or the issuer,
- a large number of shares or other investment units (especially if in smaller denominations) is offered,
- the offering is large in aggregate size, and
- the offering is conducted through a broad-based advertising campaign.[165]

This guidance gave transaction-planners and litigators some foundation for assessing whether particular offerings required registration or were exempt. Decisional law began to develop under Section 4(2), but eighteen years passed before the Supreme Court took on the issue of clarifying the nature of a public offering.

In 1953, the Court decided *SEC v. Ralston Purina Co.*, the seminal case involving the availability of the private offering exemption.[166] In *Ralston Purina*, the Court refused to impose a numerical limitation as a litmus test for determining whether a public offering has been conducted.[167] The Court's opinion indicates that the number of offerees is relevant, but not dispositive, to this determination.[168] Instead, the Court found that "the exemption question turns on the knowledge of the offerees."[169] According to the Court, "the applicability of § [4(2)] should turn on whether the

164. General Counsel Letter Regarding Section 4(1), Exchange Act Release No. 285, 111 Fed. Reg. 10952–53 (Jan. 24, 1935) [hereinafter SEC General Counsel Letter].

165. *Id.*

166. SEC v. Ralston Purina Co., 346 U.S. 119 (8th Cir. 1953).

167. *Id.* at 125. In this case, Ralston Purina claimed that an offering of treasury stock to its "key employees" was not a public offering. The group of offerees, however, included any employee who wanted to participate in the offering. *Id.* at 121. Although the Court recognized that some offerings to employees may constitute non-public offerings, the Court determined that, absent special circumstances such as an offering to certain corporate executives, "employees are just as much members of the investing 'public' as any of their neighbors in the community." *Id.* at 125–26.

168. *Id.* at 125 ("It may well be that offerings to a substantial number of persons would rarely be exempt.").

169. *Id.* at 126.

particular class of persons affected needs the protection of the Act."[170] The Court determined that where the offerees "are shown to be able to fend for themselves," they do not need the protection of the Securities Act's registration requirement, and therefore, the offering should not be characterized as a public offering for purposes of Section 4(2) of the Securities Act.[171]

Ralston Purina and its progeny have established two overarching factors that indicate whether offerees are able to fend for themselves. The first factor is the "sophistication" of the solicited investors. Offerees who possess financial and business knowledge that allows them to appreciate the risks of the investment have been considered sophisticated for these purposes.[172] Sophisticated investors can fend for themselves if they have the appropriate type and amount of information or access to it. Accordingly, the second factor is the information or access to information provided to offerees. Sophisticated offerees provided with substantially the same information—or meaningful access to substantially the same information—as that provided in a registration statement can fend for themselves.[173]

Decisional law after *Ralston Purina* clarifies that access to information is meaningful when an offeree is able to obtain the information needed to

---

170. *Id.* at 125.

171. *Id.*

172. *See* SEC v. Murphy, 626 F.2d 633, 646 (9th Cir. 1980) (in which the only evidence of investor sophistication offered was that 60% of the investors were represented by purchaser representatives, suggesting that "at least . . .the majority of the purchasers, if not the majority of offerees, lacked the sort of business acumen necessary to qualify as sophisticated investors"); Hill York Corp. v. Am. Int'l Franchises, Inc., 448 F.2d 680, 690 (4th Cir. 1971) (recognizing that the sophistication requirement was met where the offering was made "only to sophisticated businessmen and lawyers . . . ."); Lively v. Hirschfeld, 440 F.2d 631, 633 (10th Cir. 1971) ("The Supreme Court in its description of a possible 'private' group in *Ralston* Purina includes only persons of exceptional business experience . . . ."). As these and other cases illustrate, the concept of sophistication is a bit fluid. *See* C. Howard Fletcher, III, *Sophisticated Investors Under the Federal Securities Laws* 1988 DUKE L.J. 1081, 1084–85 (1988) ("[T]he federal courts' treatment of investor sophistication reflects a doctrine in disarray . . . [as] the courts' treatment of sophisticated investors shows little coherence or, if you will, reflects little cross-fertilization among the different settings in which the sophistication issue arises.").

173. The *Ralston Purina* Court only references access to information. SEC v. Ralston Purina, Co., 346 U.S. 119, 127 (8th Cir. 1953). Subsequent cases in lower courts further developed this aspect of the doctrine. *See Hill York*, 448 F.2d at 690 ("[T]he relationship between the promoters and the purchasers and the 'access to the kind of information which registration would disclose' become highly relevant factors." (citation omitted)). The offerees in *Hill York* had no previous relationship with the issuer at the time of the offering and were given only a few brochures with minimal information about the issuer. *Id.* The court concluded that the offerees "could not bring their sophisticated knowledge of business affairs to bear in deciding whether or not to invest" in the venture because they did not possess the "information requisite for a registration statement." *Id.*

make an informed investment decision. If an offeree is not actually provided information akin to that provided in a registration statement, the offeree must have access to that level of information and a relationship with the issuer that reasonably enables the offeree to take advantage of the access to ascertain the information.[174] Therefore, if the issuer does not disclose the requisite information to the offerees, the issuer must prove that its relationship with each offeree was such that it satisfied the access requirement.[175] The private offering exemption is not available unless both sophistication and disclosure of or meaningful access to information exists because "[s]ophistication is not a substitute for access to the information that registration would disclose."[176] Furthermore, disclosure of or access to important information is an empty promise without the ability to ascertain and appreciate the risks involved with the investment.[177]

By moving away from numerical limitations and focusing on the concepts of sophistication and disclosure of or meaningful access to a prescribed level of information, the opinions in *Ralston Purina* and its progeny have created "doubts and ambiguities . . . by varying Section 4(2) interpretations."[178] To clarify some of the uncertainty surrounding the application of the Section 4(2) exemption, the SEC adopted Rule 506 as part of Regulation D under the Securities Act, a safe harbor under Section 4(2).[179] We address the rule below. Considering only Section 4(2) and relevant decisional law, however, the twin concepts of sophistication and information would require crowdfunded ventures to ensure that all individuals who visit crowdfunding websites (who would then be offerees)

---

174. *See* Doran v. Petroleum Mgmt. Corp., 545 F.2d 893 (5th Cir. 1977). In *Doran*, the issuer tried to use the private placement exemption to prevent an investor from rescinding an agreement for an oil-drilling venture. *Id.* at 897. The court found that the investor, who had a degree in petroleum engineering and a net worth of over $1,000,000 (including holdings in 26 oil and gas platforms worth over $850,000), was sophisticated. *Id.* at 902. The court concluded, however, that a sophisticated investor could not have used his knowledge of business affairs to make a prudent investment decision without the information that would be contained in a registration statement. *Id.* Focusing on the information requirement, the court pointed out that where disclosure is shown, "the absence of a privileged relationship between the offeree and issuer would not preclude a finding that the offering was private." *Id.* at 904. However, when an issuer claims the offeree had access to information, "the relationship between offeree and issuer now becomes critical, for it must be shown that the offeree could realistically have been expected to take advantage of his access to ascertain the relevant information." *Id.* at 904–05.

175. *Id.* at 904 ("Such access might be afforded merely by the position of the offeree or by the issuer's promise to open appropriate files and records to the offeree as well as to answer inquiries regarding material information.").

176. *Id.* at 892 (citing United States v. Custer Channel Wing Corp., 376 F.2d 675, 678 (4th Cir. 1967)).

177. *Id.* at 904–05.

178. Cohn & Yadley, *supra* note 144, at 22.

179. *See infra* Part III.A.2.b.

meet the sophistication requirements and are given access to the required information. This is impractical because (even assuming that sophistication can be sufficiently ascertained through the Internet) the costs associated with providing the appropriate level of information to offerees over the Internet (none of whom may be assumed to have a pre-existing relationship with the issuer) are high in relation to the benefit sought, which in most cases is a relatively small amount of funding. Further, "the SEC has indicated that any 'public advertising is inconsistent with a claim of private offering.'"[180] This prohibition eliminates any hope for an open-access crowdfunding business model under the private offering exemption.[181]

b. Rules 504, 505, and 506 of Regulation D

Regulation D is a set of rules adopted by the SEC to provide exemptions principally for small issues and small issuers.[182] The main operative provisions are Rules 504 and 505,[183] adopted under the SEC's exemptive authority in Section 3(b) of the Securities Act,[184] and Rule 506,[185] adopted under Section 4(2) of the Securities Act.[186] Each exemptive rule has unique attributes, but there is some overlap in the requirements. Common to all three rules, however, are three unifying principles. First, offerings made within six months of each other may be integrated and considered to be a single offering if they have certain specified common characteristics.[187] Second, securities acquired in Regulation D offerings are considered restricted securities for purposes of the Securities Act and cannot be resold absent registration or the availability of an applicable exemption.[188] Third, except in limited circumstances under Rule 504, issuers and their agents may not offer or sell securities under Regulation D using "any form of general solicitation or general advertising."[189]

---

180. HAZEN, *supra* note 163, § 4.24 (footnote omitted).

181. *See infra* Part III.A.2.b. (discussing the prohibition of general advertising).

182. HAZEN, *supra* note 163, § 4.19[1].

183. 17 C.F.R. §§ 230.504–05 (2011).

184. 15 U.S.C. § 77c(b) (2006). Section 3(b) allows the SEC to pass rules and regulations exempting "any class of securities . . . if it finds that the enforcement of [the Securities Act] is not necessary in the public interest and for the protection of investors by reason of the small amount involved or the limited character of the public offering." *Id.* The maximum aggregate amount of any offering exempted under Section 3(b) is $5,000,000. *Id.*

185. 17 C.F.R. § 230.506.

186. 15 U.S.C. § 77d(2). Section 4(5) (formerly Section 4(6)) of the Securities Act, 15 U.S.C. § 77d(5), which allows for limited offerings to accredited investors, is also a foundation for the exemptions in Rules 505 and 506 of Regulation D. *See* HAZEN, *supra* note 163, § 4.19.

187. 17 C.F.R. § 230.502(a) (commonly referred to as "integration").

188. 17 C.F.R. § 230.502(d).

189. 17 C.F.R. § 230.502(c).

Rule 504 provides an exemption for certain offerings not exceeding an aggregate of $1,000,000 within a twelve-month period.[190] This exemption places no limits on the number of offerees or purchasers, and does not require the issuer to provide specific affirmative disclosure.[191]

Rule 505 provides an exemption for offerings with a maximum aggregate offering price of $5,000,000 within a twelve-month period.[192] The rule limits the number of purchasers to thirty-five, not including "accredited investors," a term that generally refers to entities and individuals who are presumed to be able to bear the financial risk of the total loss of their investment (e.g., institutional investors and high-net-worth individuals).[193] Additionally, Rule 505 requires disclosure of specific financial and non-financial information to any securities purchasers who are not accredited investors.[194]

Rule 506 is a Section 4(2) safe harbor included in Regulation D.[195] Because it is not based on the SEC's authority to grant exemptions under Section 3(b), Rule 506 does not limit the maximum aggregate size of an

---

190. 17 C.F.R. § 230.504(b)(2) (commonly referred to as "aggregation").

191. *See* 17 C.F.R. § 230.504(b) (apart from the exclusion of public companies, investment companies, and specified development stage companies, the only conditions that must be satisfied under this rule—other than the applicable conditions under Rule 502—relate to the $1,000,000 aggregate limitation on offering size).

192. 17 C.F.R. § 230.505(b)(2)(i). Like Rule 504, Rule 505 is not available for offerings by investment companies. *Id.* § 230.505(a).

193. 17 C.F.R. at § 230.505(b)(2)(ii). *See* Rule 501, 17 C.F.R. § 230.501(a), (e) (excluding accredited investors in calculating the number of purchasers for purposes of Rule 505 and defining "accredited investor" to generally include: banks; savings and loan associations; insurance companies; employee benefit plans; private business development companies; insiders of the issuer of the securities; any individuals whose individual net worth, or joint net worth with that person's spouse, at the time of the purchase exceeds $1,000,000; and any person with individual income of greater than $200,000, or $300,000 joint income with a spouse, in each of the two previous years).

194. 17 C.F.R. §§ 230.502(b)(1), 230.505(b)(1). Rule 505(b)(1) incorporates by reference the requirements of Rule 502, and Rule 502(b)(1) mandates the disclosure to non-accredited investors of various financial and non-financial information for offerings made under Rules 505 and 506. *Id.*

195. HAZEN, *supra* note 163, § 4.20[1]. Section 4(5) (formerly denominated Section 4(6)) of the Securities Act also covers offerings of the kind exempted under Rule 506, *see supra* note 186, but became outdated and superfluous when the more detailed safe harbor provisions of Rule 506 were adopted. *See* COX ET AL., *supra* note 141, at 286 ("Section 4(6) reflected congressional dissatisfaction with the state of limited offering exemptions in the early 1980s. The SEC responded with the adoption of Regulation D, rendering the statutory exemption of little, if any, use today."); Gary M. Brown, *Securities Act Registration Exemptions*, in 1 UNDERSTANDING THE SECURITIES LAWS, 209, 217 (Practising Law Institute 2009) ("Among other things, Regulation D incorporates the accredited investor concept of section 4(6) into a more useful exemption, making the free-standing statutory section largely superfluous.").

offering that is exempt under its provisions.[196] Like Rule 505, Rule 506 specifies that no more than thirty-five non-accredited investors may purchase the securities[197] and requires the same affirmative disclosures to all non-accredited investors.[198] Consistent with its roots in Section 4(2), Rule 506 requires that that each non-accredited investor or the non-accredited investor's "purchaser representative" meet a minimum sophistication requirement or that the issuer "reasonably believes" immediately prior to making a sale that each non-accredited purchaser or purchaser representative meets that sophistication requirement.[199]

The most serious obstacle to using Regulation D to exempt crowdfunded offerings from registration is its overall prohibition of general solicitation and advertising.[200] In fact, "[t]here is no greater impediment to the ability of small companies to raise capital under the securities laws."[201] The ban on general solicitation and advertising is a substantial obstacle because the SEC has interpreted this restriction very broadly by construing "'general solicitation' to include offers to any person with whom the issuer, or the issuer's agent, has not had a prior relationship."[202] Rules 505 and 506

196. *See supra* note 184 and accompanying text.

197. 17 C.F.R. § 230.506(b)(2)(i); *see also supra* note 193 and accompanying text (regarding the parallel requirement in Rule 505).

198. 17 C.F.R. §§ 230.502(b), 230.506(b)(1); *see also supra* note 194 and accompanying text (regarding the parallel requirement in Rule 505).

199. 17 C.F.R. § 230.506(b)(2)(ii) (requiring that each non-accredited investor, or his or her purchaser representative, have "such knowledge and experience in financial and business matters that he is capable of evaluating the merits and risks of the prospective investment," or that the issuer "reasonably believe" that the purchaser meets that requirement).

200. *See* Cohn & Yadley, *supra* note 144, at 11 ("The SEC's ban on general advertising and general solicitation in private offerings . . . eliminates the potential of the internet to attract investors . . . .").

201. *Id.* at 36; *see also* sources cited *infra* note 357 (arguing for dismantlement of the ban on general solicitation and general advertising). This prohibition also applies to the private offering exemption under Section 4(2), discussed *supra* Part III.A.2.a. The only transactional exemptions that are not subject to this restriction are the intrastate offering exemption under Section 3(a)(11), 15 U.S.C. § 77c(a)(11) (2006), Rule 147, 17 C.F.R. § 230.147 (2011), and Regulation A, 17 C.F.R. §§ 230.251-263, discussed *infra* Part III.A.2.c.

202. Cohn & Yadley, *supra* note 144, at 41 (citations omitted). In examining a case where offerors engaged in a general solicitation by sending materials to an unknown number of people with whom the offerors did not have a pre-existing relationship, the SEC stated:

> These persons were selected only because their names were on lists that were purchased or created by Kenman. Although the make-up of the lists may indicate that the persons themselves have some degree of sophistication or financial well-being, utilization of lists of thousands of persons with no pre-existing relationship to the offeror clearly does not comply with the limitations of Rule 502(c) on the manner of solicitation.

In re Kenman Corp., Exchange Act Release No. 34-21962, 32 SEC Docket 1352-1 (Apr. 19,

completely prohibit general solicitation and advertising.[203] Rule 504 permits this manner of offering only if the issuer complies with applicable state law that provides adequate investor protection.[204]

In almost all cases,[205] crowdfunding websites exist to invite the general public—the crowd—to help fund small business ventures.[206] Their purpose is to allow ventures to access capital that they would not be able to access without using the crowdfunding website. If ventures seeking funding were able to raise the necessary capital from those with whom they have a prior relationship, they have no need for crowdfunding. Thus, the nature of crowdfunding requires the use of general solicitation and advertising. Screening devices, such as password-protected access to the crowdfunding website, are impractical in this environment. Accordingly, the typical crowdfunded venture is precluded from using the exemptions under Rules 505 and 506. Exemption under Rule 504 may be available, but only if the issuer's offering meets the state law exemption requirements set forth in Rule 504(b)(1).[207] Assuring compliance with Rule 504 for a crowdfunded venture is not straightforward. It may be difficult to determine the states in which crowdfunding interests are offered and sold, and assuming that the applicable state laws meet the proper threshold level of investor protection,

---

1985).

203. 17 C.F.R. §§ 230.502(c), 505(b)(1), 506(b)(1).

204. 17 C.F.R. § 230.504(b)(1). Rule 504(b)(1) requires that the issuer comply with state laws compelling public filing or delivery of disclosure documents before the sale of securities or that the securities be sold exclusively according to state law exemptions that allow for general solicitation to accredited investors. *Id.*

205. When this article was in draft form, we noted that one crowdfunding website, PROFOUNDER, only allowed investors who were invited by the small business owner seeking the funding. In that case, the crowdfunding website's primary function is to market the small business's products and performing administrative work. Although the website facilitates fundraising by giving the small business owner a platform from which to promote the business, ProFounder's fundraising strategy is not based on allowing the small business owner to access more sources of capital; it is based on giving the small business owner a more effective strategy for tapping those resources. *For Entrepreneurs: FAQs*, ProFounder, www.profounder.com/entrepreneurs/faqs (last visited Sept. 1, 2011). The ProFounder principals were forced to redesign the ProFounder business model after a recent cease and desist order from the California Department of Corporations regarding ProFounder's status as an unlicensed broker dealer in the state. *See* Angus Loten, *Crowd-Funding Brings Unease*, WALL ST. J. ONLINE, Nov. 17, 2011, http://online.wsj.com/article/SB10001424052970203611404577042333598282986.html?mod=WSJ_SmallBusiness_LEADNewsCollection (last visited Dec. 26, 2011). The current version of the website has more of an educational and supportive mission. *See, e.g.*, *Frequently Asked Questions*, PROFOUNDER, https://www.profounder.com/faq (last visited Dec. 26, 2011).

206. *See* Belleflamme et al., *supra* note 3, at 5 ("Raising funds by tapping a general public (or the crowd) is the most important element of crowdfunding. This means that consumers can volunteer to provide input to the development of the product, in this case in form of financial help.").

207. *See* 17 C.F.R. § 230.504(b)(1).

the cost of complying with multiple state laws could be high, if not prohibitive.

The ban on general solicitation and advertising is a veritable showstopper for ventures contemplating the use of a Regulation D exemption for crowdfunding or other Internet offers and sales.[208] However, the ban is not the only obstacle that issuers face in exempting crowdfunded offerings from registration under Regulation D. The thirty-five purchaser limit under Rules 505 and 506 also is an impediment. Information available on crowdfunding websites leaves open the possibility that each venture will be funded through the purchase of interests by more than thirty-five entities and individuals.[209] We can safely assume that many, if not most, purchasers of crowdfunding interests are not accredited investors (i.e., many or most are neither institutional investors nor high-net-worth individuals). If more than thirty-fivedistinct non-accredited investors purchase crowdfunding interests in a particular venture, the offering would not qualify for an exemption under Rule 505 or Rule 506.[210] Even if fewer than thirty-five non-accredited investors were to acquire crowdfunding interests in an offering meeting the general solicitation and advertising requirements, preparation of the disclosure documents required for non-accredited investors under Rules 505 and 506 likely would be cost-prohibitive.[211]

An additional hurdle exists with respect to complying with the sophisticated investor requirements in Rule 506. It is unlikely that all purchasers of crowdfunding interests—as members of an undifferentiated Internet-based crowd—would meet the sophistication standards or have access to a qualified purchaser representative.

For these reasons, Rules 504, 505, and 506 are ill suited to exempt small-dollar-value Internet offerings to the masses from the registration requirements of Section 5 of the Securities Act. Regulation D fails to provide a viable exemption option for crowdfunded businesses.

---

208. *See, e.g.*, Press Release, U.S. Sec. & Exch. Comm'n, *SEC Enters Cease and Desist Order in Connection with Online Campaign to Buy Beer Company*, June 8. 2011, *available at* http://www.sec.gov/news/press/2011/2011-122.htm (describing alleged violations of Section 5(c) of the Securities Act in connection with the online solicitation of pledges for up to $300 million to purchase the Pabst Brewing Company).

209. For example, a crowdfunded venture that has a target goal of $15,000 and provides information on rewards for contributions in $10, $30, $50, and $100 increments may require as many as 1,500 investors to reach the funding target. If an individual was allowed to contribute less than $10 (for example, if the minimum contribution is $1), it is possible that there could be as many as 15,000 investors involved in reaching the venture's funding target.

210. *See supra* note 197 and accompanying text.

211. *See* 17 C.F.R. § 230.502(b)(2). In some instances, the disclosure may require the same kind of financial non-financial information contained in a registration statement.

c. Regulation A

Regulation A exempts offerings not exceeding $5,000,000 within a twelve-month period, provided that the issuer offers the securities using an offering statement on Form 1-A, a filing similar limited registration statement, and an offering circular, a disclosure and selling document similar to the traditional Section 10 prospectus that forms a part of the registration statement in offerings registered under the Securities Act.[212] The offering circular is a "rather full disclosure document . . . complete with financial statements prepared in accordance with generally accepted accounting principles, filed and reviewed by the SEC in a manner similar to the filing and review of registration statements."[213] Absent from Regulation A, however, are the prohibitions against general solicitation and advertising, limitations on the number of investors, and investor qualification standards.[214]

Although Regulation A alleviates many of the burdens imposed by Regulation D and has been used for at least one early Internet-based direct public offering,[215] the expense of producing the offering circular, in addition to the costs associated with state securities law compliance, makes this exemption too costly for many crowdfunded ventures.[216] Therefore, Regulation A fails to provide a practical exemption from federal securities laws for crowdfunded ventures.

---

212. 17 C.F.R. §§ 230.252–253; HAZEN, *supra* note 163, § 4.17[1].

213. Cohn & Yadley, *supra* note 144, at 28 (citations omitted); *see also* Campbell, *supra* note 146, at 105 ("Although less extensive than the corresponding disclosures required in a prospectus in a registered offering, the narrative disclosures in an offering circular are substantial." (footnote omitted)).

214. 17 C.F.R. § 230.251. Rule 251 acknowledges that the exempted offering is a "public offer or sale of securities." *Id.* Therefore, there is no need to limit general solicitation and advertising, assess accredited investor status or examine investors' sophistication because the exemption presupposes a public offering.

215. *See* Nikki D. Pope, *Crowdfunding Microstartups: It's Time For The Securities And Exchange Commission To Approve A Small Offering Exemption*, 13 U. PA. J. BUS. L. 101 (2011).

216. *See* Campbell, *supra* note 146, at 105–10. In 1997, the average cost of a Regulation A offering was $40,000 to $60,000, and the average cost of a registered offering using Form S-1 was between $400,000 and $1,000,000. HAZEN, *supra* note 163, § 4.17[1]. Although mini-registration under Regulation A costs less than a registered offering, the expense of a Regulation A offering will often still be more than the amount of capital that the crowdfunding venture seeks to raise. Thus, "the Regulation A procedure 'has for the most part become too cumbersome and expensive for small financings in an enterprise's early years.'" Cohn & Yadley, *supra* note 144, at 28 (quoting Julian M. Meer, *The Private Offering Exemption Under the Federal Securities Act – A Study in Administrative and Judicial Contraction*, 20 SW. L.J. 503, 504 (1966)).

3. Whose Conduct is Regulated?

The analysis in the two preceding subparts of Part III.A does not directly address the question of who, in a crowdfunded offering, is subject to the registration requirements of Section 5 of the Securities Act.[217] It is important to address this part of the regulatory equation as a predicate to an evaluation of both the benefits and burdens of U.S. securities regulation in the crowdfunding context and the desirability and efficacy of any adjustments that may be made to the existing regulatory framework to better serve the policy objectives applicable to securities offerings under the Securities Act. This subpart engages that analysis as it relates to the registration requirement under the Securities Act.[218]

As earlier noted, Section 5 regulates the offer and sale of a security.[219] Different categories of persons with different roles in securities transactions are recognized under the 1933 Act as persons who may offer or sell securities. Paramount among them is the issuer. As a general matter, the issuer must register securities for offer and sale.[220] Section 2 of the 1933 Act defines an issuer as "every person who issues or proposes to issue any security . . . ."[221] Under this vague definition, either the crowdfunded venture or a crowdfunding website that promotes a crowdfunded venture could be an issuer. In *SEC v. Murphy*,[222] the Ninth Circuit offered that, for purposes of determining the issuer of securities in a limited partnership, the issuer was the "entity about which the investors needed information."[223] The court limited its guidance in *Murphy* to situations involving limited partnerships and left for another day the issue of whether additional individuals or entities with information material to the investment decision would be a securities issuer.[224] This definitional guidance may best support

217. The analysis also does not address the potential effects of applicable liability provisions under the Securities Act and the Exchange Act.

218. In limiting our analysis here to matters under the Securities Act, we recognize that we fail to address other important potential roles that crowdfunding websites may occupy in securities transactions. *See* Bradford, *supra* note 4, at 32–51 (analyzing crowdfunding websites as potential exchanges, brokers, or investment advisors); Hazen, *supra* note 7, at 15–16 (analyzing the status of crowdfunding intermediaries); Bradford, *Peer-to-Peer Lending*, *supra* note 7 ("If the sites are offering securities, the sites themselves could be brokers, or even exchanges, within the meaning of the Securities Exchange Act. Alternatively, it is at least possible that crowd-funding sites are investment advisers subject to regulation under the Investment Advisers Act.").

219. 15 U.S.C. § 77e (2006).

220. 15 U.S.C. § 77f(a) ("Any security may be registered with the Commission . . . by filing a registration statement in triplicate, at least one of which shall be signed by *each issuer* . . . ." (emphasis added)).

221. 15 U.S.C. § 77b(a)(4).

222. 626 F.2d 633 (9th Cir. 1980).

223. *Id.* at 643–44.

224. *Id.* at 644.

labeling the specific crowdfunded venture as the issuer. That result seems intuitively correct, because a profit-seeking crowdfunder needs information about the venture being funded to assess the desirability and financial promise of an investment in that venture.

The conclusion that specific crowdfunded ventures are Securities Act issuers, however, does not foreclose the conclusion that crowdfunding websites also may be issuers. In fact, decisional law explicitly recognizes the possibility that multiple entities may act as "co-issuers."[225] For example, the concept of co-issuers was implicitly recognized in *Howey*, where the Court found that two affiliated companies violated Section 5 by offering a land sales contract and a related service agreement for that land.[226] Co-issuer status among affiliates was also implicitly recognized in *SEC v. Edwards*, a case involving a corporation and its subsidiary that sold payphones and offered a five-year leaseback and management agreement in conjunction with the sales.[227] Thus, where affiliated companies act together to offer or sell an investment contract that constitutes a security, both entities will likely be considered co-issuers of those securities.

However, in many crowdfunding arrangements, the crowdfunded venture and the crowdfunding website are not affiliated. This should not make a difference in whether the two are offering and selling securities for purposes of the Securities Act. The D.C. Circuit's opinion in *SEC v. Life Partners, Inc.*[228] provides limited support for the proposition that two unaffiliated entities working together to offer a security will be similarly regulated for purposes of the Securities Act's registration requirements. In *Life Partners*, Life Partners, Inc. ("LPI") offered viatical settlements[229] to investors and, along with Sterling Trust Company, an independent escrow agent acting for LPI, performed post-purchase administrative functions to ensure that investors collected on the settlements.[230] The court determined that all prongs of the *Howey* test other than the "efforts of others" prong

225. SEC v. Datronics Eng'rs, Inc., 490 F.2d 250, 254 (4th Cir. 1973), *cert. denied*, 416 U.S. 937 (1974). Datronics was an engineering company that spun off unregistered shares of nine new merger-corporations. *Id.* at 252–53. The court determined that "[c]learly, in these transactions the merger-corporation was an issuer; Datronics was a purchaser as well as a co-issuer . . . ." *Id.* at 254. In fact, the 1933 Act definition expressly recognizes the possibility of more than one issuer. *See supra* note 221 and accompanying text.

226. SEC v. W. J. Howey Co., 328 U.S. 293, 299 (1946); *see also supra* note 28 and accompanying text.

227. SEC v. Edwards, 540 U.S. 389, 391 (2004).

228. SEC v. Life Partners, Inc., 87 F.3d 536 (D.C. Cir. 1996).

229. Viatical settlements are financial arrangements through which an investor purchases an interest in the life insurance policy of a terminally ill individual (typically an AIDS patient) at a twenty to forty percent discount. *Id.* at 537. This arrangement provides the patient with cash, and the investor's profit is the difference between the discounted price paid for the policy and the death benefit collected from the insurer. *Id.*

230. *Id.* at 540.

were satisfied.[231] An important aspect of this decision as it relates to those regulated under Section 5 of the Securities Act is that the court analyzed the efforts of two unaffiliated entities together.[232] Specifically, the court recognized that two separate entities can work together in such a manner that their functions are inseparable for purposes of determining whether a security is being offered. Because investors would need information about both parties, each must comply with the Securities Act's registration requirements. Both parties, even if unaffiliated, should be treated as co-issuers where their efforts are inseparable and integral to the offering or sale of the contract, transaction, or scheme that constitutes a security under the Securities Act. Accordingly, in some circumstances, crowdfunded ventures and crowdfunding websites may be acting in concert in a manner that makes them co-issuers.

An alternative to characterizing the crowdfunded venture and the crowdfunding website as co-issuers is to treat the crowdfunded venture as an issuer and the crowdfunding website as an underwriter. "Underwriter status subjects applicable parties to the provisions of Section 5 and results in liability exposure for material misrepresentations and nondisclosures contained in the registration statement."[233] The Securities Act includes a definition of "underwriter":

> The term "underwriter" means any person who has purchased from an issuer with a view to, or offers or sells for an issuer in connection with, the distribution of any security, or participates or has a direct or indirect participation in any such undertaking, or participates or has a participation in the direct or indirect underwriting of any such undertaking . . . .[234]

This definition encompasses any party "who offers or sells for an issuer in connection with a distribution."[235] Courts have generally followed the Second Circuit's reasoning in *Gilligan, Will & Co. v. SEC*, which states that a distribution is synonymous with a "public offering," as that term is defined in *Ralston Purina*.[236] In other words, if investors in an offering need the protection of the Securities Act, the offering is a distribution. Someone who offers or sells securities for an issuer in that offering, or who participates in that offer or sale, is an underwriter under the Securities Act. A common function of crowdfunding websites is to provide each featured crowdfunded venture with access to a base of investors that may be willing

231. *Id.* at 542–45.

232. *Id.* at 546.

233. MARC I. STEINBERG, UNDERSTANDING SECURITIES LAWS 175–76 (5th ed. 2009).

234. 15 U.S.C. § 77b(a)(11) (2006).

235. COX ET AL., *supra* note 141, at 339.

236. *See* Gilligan, Will & Co. v. SEC, 267 F.2d 461, 466–67 (determining that the "public offering" issue is "dispositive of the question whether petitioners 'purchased . . . with a view to . . . distribution'").

to purchase interests in that venture and to assist in promoting the venture to those potential investors. In this sense, crowdfunding websites perform much the same traditional function that investment banks play when they serve as underwriters in prototypical underwritten public offerings: the identification of potential investors and the promotion of the issuer and the securities being offered and sold.[237] These crowdfunding websites apparently offer or sell securities for an issuer—the crowdfunded venture—in that offering (or at least participate in that offer or sale).

Of course, investment banks serving as underwriters in public offerings typically act as conduits for securities distribution by purchasing the securities from the issuer and reselling them to the public. But activity as a conduit for the securities is not required; activity as a promoter is sufficient to establish underwriter status. In *SEC v. Chinese Consolidated Benevolent Ass'n*,[238] the defendant association merely solicited purchase orders for the securities and engaged in limited ministerial activities in connection with the sale of Chinese government bonds, yet the court ruled that the defendant was an underwriter. The court referenced the relevant language from the Securities Act:

> [T]he words "[sell] for an issuer in connection with the distribution of any security" ought to be read as covering continual solicitations, such as the defendant was engaged in, which normally would result in a distribution of issues of unregistered securities within the United States. Here a series of events were set in motion by the solicitation of offers to buy which culminated in a distribution that was initiated by the defendant. We hold that the defendant acted as an underwriter.[239]

The offering-related tasks that the association undertook in *Chinese Consolidated Benevolent Ass'n* were similar to tasks that crowdfunding websites perform. As an underwriter, the crowdfunding website could be liable for Section 5 violations, even if it is not a co-issuer.[240]

---

237. We note that people who perform these functions also may be deemed finders, who may be classified as brokers for purposes of the Exchange Act. *See* Bradford, *supra* note 4, at 33–43; Hazen, *supra* note 7, at 16; Brumberg, Mackey & Weil, P.L.C., No-Action Letter (May 17, 2010), *available at* http://www.sec.gov/divisions/marketreg/mr-noaction/2010/brumbergmackey051710.pdf; *see also* John L. Orcutt, *Improving the Efficiency of the Angel Finance Market: A Proposal to Expand the Intermediary Role of Finders in the Private Capital Raising Setting*, 37 ARIZ. ST. L.J. 861, 897–920 (2005) (describing the role and legal status of finders in securities offerings). We also note that crowdfunding websites may perform additional functions (including post-sale administrative functions) that are not easily classified as underwriting or finding services. *See, e.g.*, text accompanying *supra* note 230 (describing this kind of activity in the *Life Partners* case).

238. 120 F.2d 738 (2d Cir. 1941).

239. *Id.*

240. *Accord* Hazen, *supra* note 7, at 15–16.

The doctrine of participant liability, as fashioned by judicial decisions, supports regulation of the conduct of both the crowdfunded venture and the crowdfunding website under the Securities Act as offerors or sellers of securities. Under this doctrine, Section 5 liability attaches to an individual or entity that has a "significant role" in the offer or sale of securities.[241] A "significant role" has been defined to "include one who is both a 'necessary participant' and a 'substantial factor' in the sales transaction."[242] Where an offering participant is both a necessary participant and a substantial factor in an offering that violates Section 5, that participant is liable for the violation. This liability apparently is a form of underwriter liability, because an individual or entity who has a "significant role" in an offering is a participant in a distribution of securities by an issuer and, therefore, an underwriter.[243] Under participant liability, there is no question that Section 5 compliance and liability will attach to both the crowdfunded venture and the crowdfunding website. Both are necessary participants in the offering; there would be no offer or sale of a crowdfunding interest without the crowdfunded venture, and the crowdfunding website is the essential vehicle for the offer and sale of the crowdfunding interests. Moreover, crowdfunding websites may be involved in the distribution of the crowdfunding interests offered on their sites in many ways other than as a fundraising host site for crowdfunded ventures. Examples of this involvement include: screening the projects prior to offering the investment to the public,[244] making promotional videos or designing individual web pages for the crowdfunded ventures,[245] serving as conduits for invested funds,[246] collecting cash from the crowdfunded venture and distributing it to investors in accordance with the profit-sharing or revenue-sharing components of the crowdfunding interests promised to investors at the time

241. SEC v. Phan, 500 F.3d 895, 906 (quoting SEC v. Murphy, 626 F.2d 633, 652 (9th Cir. 1980)).

242. *Id.*

243. *See* SEC v. Allison, No. C-81-19 RPA, 1982 WL 1322, at *3 (N.D. Cal. Aug. 11, 1982) ("When, as in this case, a defendant's actions were necessary to and a substantial factor in an illegal securities distribution, the defendant is a participant and thus an underwriter irrespective of the defendant's intent." (citing *Murphy*, 626 F.2d at 648–50)).

244. For example, MicroVentures pre-screens all business ideas to evaluate whether the business meets the criteria to be listed on the site. *See Investors: Frequently Asked Questions*, MICROVENTURES, http://www.microventures.com/investors/faq (last visited Dec. 26, 2011). Some factors include the company's business plan, business experience of the company's leaders, how the business will use the funds, and the risk to investors. *Id.*

245. ProFounder, for example, formerly created individual fundraising websites to market the business ventures that were selected for funding. *For Investors: How it Works*, PROFOUNDER, http://www.profounder.com/entrepreneurs (last visited Sept. 1, 2011).

246. Microfinancing enterprises, such as Kiva, match investors' funds to microfinancing institutions in less developed countries, which provide funding to the principal. *How Kiva Works*, www.kiva.org/about/how (last visited Dec. 26, 2011).

the interests were purchased,[247] and performing accounting and other ministerial functions for crowdfunded ventures.[248] Both the crowdfunded venture and the crowdfunding website are substantial factors in the sales transaction because both work together to provide potential and actual investors with a financial interest in the underlying venture. In the prototypical crowdfunding model, the principals behind the crowdfunded venture create and manage the business or project needing funds while the crowdfunding website attracts investors, supports them in their chosen investments, and administers the relationship between the crowdfunded venture and its investors.[249] The doctrine of participant liability supports imposing the requirements of Section 5 of the Securities Act on both crowdfunded ventures and crowdfunding websites.

Finally, we note the possibility that crowdfunding websites could be brokers, investment advisors, exchanges, another form of intermediary, or investor fiduciaries under federal or state securities laws.[250] We have left the analysis of those possibilities to others.[251]

B. Policy Underpinnings

The system of regulation described above serves two overarching policies: protecting investors and maintaining market integrity.[252] These policies are effectuated principally through doctrinal rules that provide for mandatory disclosure and liability for noncompliance, material misstatements and omissions, and fraud. However, Congress and the SEC also have used substantive regulation of constituents and conduct to

247. Appbackr facilitates the development and sale of newly created mobile apps on the Apple App Store. Appbackr, like most crowdfunding sites with revenue-sharing or profit-sharing arrangements, acts as the intermediary between funders and principals. *See How Does it Work?*, APPBACKR, http://www.appbackr.com/static/learnMore (last visited Dec. 26, 2011) ("We act as the intermediary between developers and backers.").

248. Fansnextdoor, which seeks investments from fans to fund creative ventures, retains 3% of the funds raised, for which it provides "webmastering services, web maintenance, rewriting and translations [EN-FR] when it is necessary." *What is fansnextdoor?*, FANSNEXTDOOR, http://en.fansnextdoor.com/help/how-it-works (last visited Dec. 26, 2011). Additionally, Fansnextdoor states that "[w]e . . . also . . . adjust at a certain level the steps of project creators, and . . . help you communicate during and after the project . . . ." *Id.*

249. *See supra* notes 244–48 and accompanying text.

250. *See supra* notes 7, 218, 237.

251. *See, e.g.*, Bradford, *supra* note 4, at 32–51 (discussing the status of crowdfunding websites under federal securities laws).

252. *See* LOUIS LOSS & JOEL SELIGMAN, FUNDAMENTALS OF SECURITIES REGULATION 385 (5th ed. 2010); *see also* Stephen J. Choi & Andrew T. Guzman, *Portable Reciprocity: Rethinking the International Reach of Securities Regulation*, 71 S. CAL. L. REV. 903, 941–44 (1998).

effectuate investor protection and maintain market integrity, particularly since the adoption of the Sarbanes-Oxley Act of 2002.[253]

Registration is the vehicle for mandatory disclosure under the Securities Act. Liability results from a failure to register offers or sales of securities and from fraudulent or other objectionable activities (e.g., material misstatements or omissions in registration statements and prospectuses) in connection with the registration requirement.[254] Congressional and SEC rulemaking and decision-making under the Securities Act focuses on supporting investor protection and market integrity in this context. For example, the SEC's general exemptive authority under the Securities Act is subject to the requirement that the exemption be "necessary or appropriate in the public interest, and . . . consistent with the protection of investors."[255] The Securities Act also imposes market-oriented requirements on the SEC's rule-making authority.[256]

The various types of statutory and regulatory exemptions under the Securities Act are rooted in different subsidiary policies consistent with the protection of investors and markets. For example, as the Court recognized in the *Ralston Purina* case,[257] it appears that Congress intended Section 4(2) to allow for offerings that are limited in character in a way that makes the Securities Act's registration and liability protections unnecessary for maintaining adequate investor and market protections:

---

253. *See* Ernst & Ernst v. Hochfelder, 425 U.S. 185, 195 (1976) ("The Securities Act . . . was designed to provide investors with full disclosure of material information concerning public offerings of securities in commerce, to protect investors against fraud and, through the imposition of specified civil liabilities, to promote ethical standards of honesty and fair dealing."); Susanna Kim Ripken, *The Dangers and Drawbacks of the Disclosure Antidote: Toward a More Substantive Approach to Securities Regulation*, 58 BAYLOR L. REV. 139, 142–43 (2006) (describing substantive regulation under Sarbanes-Oxley). Mandatory disclosure serves to inform investor decision-making and enhance the efficiency of the market. *See* HAZEN, *supra* note 163, at 168–69; Ripken, *supra*, at 145 ("For the last seventy years, federal securities legislation in general has consistently relied on the philosophy of disclosure as the primary tool for protecting investors and regulating the securities market.").

254. *See supra* notes 137–42.

255. 15 U.S.C. § 77z-3 (2006).

256. 15 U.S.C. § 77b(b).

> Whenever pursuant to this subchapter the Commission is engaged in rulemaking and is required to consider or determine whether an action is necessary or appropriate in the public interest, the Commission shall also consider, in addition to the protection of investors, whether the action will promote efficiency, competition, and capital formation.

Id.

257. *See supra* notes 170–71 and accompanying text.

> The legislative history is of little help except insofar as the general tone may be set by the House Committee's reference to this exemption as permitting "an issuer to make a specific or an isolated sale of its securities to a particular person," and to the exemption generally as directed to transaction "where there is no practical need for [the bill's] application or where the public benefits are too remote."[258]

Section 3(b) is founded on different objectives, however. It seeks to encourage capital formation through small offerings likely to be used to finance small ventures:

> A perennial conundrum of the securities laws is how to treat small businesses fairly. Historically, considerable evidence has demonstrated that a substantial proportion of securities fraud is committed by the promoters of new, speculative firms. To fully exempt small business from the reach of the securities laws would deprive investors of protection in some of the instances where investors need protection most. On the other hand, there is no question that when small firms issue new securities they pay a proportionately higher price for underwriting compensation (the primary expense), accounting, legal, and filing costs than larger businesses. For some small firms, the costs of a public securities distribution are prohibitive. Unless it is national policy to give the large business firms advantages over the small in capital formation, it is essential to create compensatory programs to stimulate the financing of small firms.[259]

By its express terms, Section 3(b) is constrained by the overall policy aims of the Securities Act. Section 3(b) only authorizes the SEC to adopt exemptions for offerings with an aggregate value of $5,000,000 or less, and only "if it finds that the enforcement of this title [the Securities Act] with respect to such securities is not necessary in the public interest and for the protection of investors by reason of the small amount involved or the limited character of the public offering."[260] Regulation A is an example of SEC rulemaking that effectuates these purposes. Regulation D, which relies on both Section 4(2) and Section 3(b) of the Securities Act for its statutory authority, encompasses the underlying policies of both sections by efficiently combining and enhancing earlier registration exemptions adopted under Sections 3 and 4 of the Securities Act.[261] Where does crowdfunding fit in?

Crowdfunded offerings, as currently conducted, are not private offerings; by their nature, crowdfunded offerings are not limited offerings (in terms of their ability to reach potential investors) and are not isolated

258. LOSS & SELIGMAN, *supra* note 252, at 395.

259. *Id.* at 387.

260. 15 U.S.C. § 77c(b).

261. HAZEN, *supra* note 163, at 418–19.

offers and sales of securities. There may be no practical need for application of the Securities Act's registration scheme or the registration regime's public benefits may be too remote for crowdfunded offerings of a relatively small number of units or offerings that are small in aggregate dollar value. In other words, registration of crowdfunded offerings may not be necessary in the public interest and for the protection of investors. Crowdfunded offerings typically are small in size—with aggregate offering values significantly lower than the $5,000,000 maximum for SEC exemptions under Section 3(b) of the Securities Act.[262] Accordingly, considering that the existing regulatory framework is unfriendly to crowdfunded offerings and assuming that crowdfunding is an activity that is desirable and consistent with the policies of investor protection and maintenance of market integrity, it seems appropriate to consider an exemption from registration for crowdfunded offerings.

IV. EXEMPTING CROWDFUNDING OFFERINGS FROM SECURITIES ACT REGISTRATION REQUIREMENTS

Positing that crowdfunding may be desirable and that a registration exemption may be appropriate, this Part first articulates an argument for providing a registration exemption for crowdfunding and then proceeds to propose elements of a possible crowdfunding exemption.

A. Why Consider an Exemption for Crowdfunding?

Crowdfunding is a very new corporate finance tool, and existing applicable law shapes (even when it does not constrain) crowdfunding's current parameters. Accordingly, a comprehensive analysis of crowdfunding's current and projected future benefits and costs is not yet possible. Nevertheless, it is feasible to review some of crowdfunding's perceived and actual advantages and disadvantages to venturers, investors, other market participants, and the public at large.[263]

---

262. Not much empirical data yet is available on crowdfunded ventures and offerings. However, a recent study by two Belgian business scholars showed that the mean amount of funds raised by the 33 crowdfunded ventures for which finding data was available was $3.5 million, and the median amount of funds raised was under $29,000. Belleflamme et al., *supra* note 3, at 17 (Table 2).

263. Because "[c]rowdfunding is a market of and for the participants," some traditional financial intermediaries may be shut out of this sector of the capital formation process. LAWTON & MAROM, *supra* note 4, at 162. No doubt, however, new support roles for crowdfunding will develop as the industry matures. Crowdfunding sites like 33needs are new forms of financial intermediation. *See* Belleflamme et al., *supra* note 3, at 4. We make no attempt here to assess the various social and economic tradeoffs among transaction participants that inevitably will occur as crowdfunding further develops. *See generally* LAWTON & MAROM, *supra* note 4, at 167–72 (describing a few ways in which traditional

1. Perceived and Actual Advantages of Crowdfunding

Crowdfunding may solve a key problem that small businesses have in funding their operations: locating a sufficient number of potential and actual investors in a cost-effective manner. Most people seeking to fund businesses and projects, especially younger entrepreneurs, do not have relationships with enough entities and individuals to create a stable source of venture capital without third-party assistance.[264] In addition, these same venturers often have few connections to people who can find investors for them (and even if they do have these connections, they are unlikely to have the funds necessary to retain these individuals and access their services).[265] The Internet has made locating investors much more efficient:

> The scalability of classic human-centric networks has hit the skids. Fortunately, while the Internet has to a large degree exacerbated this problem, it also holds many solutions. We live in the age where a couple billion of people use the Internet, and social networking has become part of our lives, whether it be using Facebook, Twitter, LinkedIn, Foursquare, Blippy, Quora, YouTube, blogging or otherwise. The irony is, many individuals who have created big social networking presences have a bigger "Rolodex" than many financiers.[266]

Crowdfunding enables entrepreneurs to more quickly and easily identify supporter-investors who are willing and able to fund their businesses or projects. These investors may be more likely to be engaged with, and even passionate about, the ventures they are funding than repeat players in the seed, angel, or venture capital game. Many of these investors are not otherwise involved in funding business ventures and were an untapped source of small business capital prior to the advent of crowdfunding. Crowdfunding gives these investors a way to participate in corporate finance that they may not otherwise have. Specifically, crowdfunding provides a new outlet for the capital of ardent consumer-

---

finance professionals could engage with crowdfunding). That analysis must wait for another day.

264. *See* Campbell, *supra* note 146, at 89 ("Usually, company employees do not know where to find potential investors"). *Cf. id.* at 81 ("[S]mall businesses face daunting economic and structural conditions when they enter the capital markets. External capital for them is hard to find and expensive to acquire.").

265. *Id* at 81. ("The absence of financial intermediation services for small businesses means that they are almost always on their own to find investors; their small capital needs mean that their relative offering costs are often sky high.").

266. LAWTON & MAROM, *supra* note 4, at 55; *see also* Belleflamme et al., *supra* note 3, at 6 ("[T]he development of Web 2.0 is a critical ingredient that has facilitated the access to the 'crowd.' Roughly speaking, Web 2.0 is a Web-as-participation-platform that facilitates interaction between users. This structure is crucial for entrepreneurs to be able to easily reach networks of investors or consumers." (footnote omitted)).

investors on the Internet, where these potential funders spend much of their time. Because of their particularized interest in the ventures they choose to fund and their Internet savvy, investors in crowdfunded ventures may choose to fund businesses and projects different from those funded through more traditional capital-raising methods.[267] These crowdfunded ventures may be more welfare-enhancing or more successful in their relevant product or service markets than businesses and projects funded through standard venture capital financings.[268]

Crowdfunding also has the potential to help stimulate the economy through the efficient financing it provides to some small businesses. Small businesses have the capacity to be an engine of economic growth by creating jobs—and providing hope.[269] However, the difficulty that small businesses have in funding their operations has worsened as a result of the current economic crisis.[270] Crowdfunding may help generate the capital small businesses need to commence or continue operations, which in turn, fuels economic growth. For example, veteran crowdfunding site IndieGoGo is among the participants in Startup America,[271] a White House initiative

267. *See* Belleflamme et al., *supra* note 3, at 28 ("Compared to other means of financings, crowdfunding opportunities exhibit several important differences that are likely to affect risk-return profile of investors and motivations for providing money to crowd-funders.").

268. *See* LAWTON & MAROM, *supra* note 4, at 55 ("[W]ith the hyper-awareness and immersion that comes from using these modern [social networking] tools, many individuals in the crowd have a much better chance of screening and picking the best and most interesting new projects."); Belleflamme et al., *supra* note 3, at 28 (Crowdfunding "is a unique way to validate original ideas in front of a specifically targeted audience. This may in turn provide insights into market potential of the product or service offered.").

269. *See* Campbell, *supra* note 146, at 81 (footnote omitted) ("Society needs small businesses. They are vital to our national economy, both qualitatively and quantitatively. They account for as much as 40% of our total economic activity and provide consumers with many of the services and products that are essential in our day-to-day lives."); *id.* at 84–86 (chronicling the social and economic importance of small business); Orcutt, *supra* note 237, at 861–62 (referencing information from the U.S. Small Business Administration Office of Advocacy); William K. Sjostrom, Jr., *Relaxing The Ban: It's Time To Allow General Solicitation And Advertising In Exempt Offerings*, 32 FLA. ST. U. L. REV. 1, 1 (2004) ("Small businesses play a pivotal role in the United States economy. 'They are the foundation of the Nation's economic growth: virtually all of the new jobs, 53 percent of employment, 51 percent of private sector output, and a disproportionate share of innovations come from small firms.'" (footnote omitted)); Katherine Reynolds Lewis, *Crowdfunding Promoted to Help Small Businesses*, FISCAL TIMES (Apr. 17, 2011), http://www.thefiscaltimes.com/Articles/2011/04/17/Crowdfunding-Promoted-to-Help-Small-Businesses.aspx ("President Obama launched Startup America to encourage entrepreneurship, stressing that small businesses traditionally have been the engine of job creation, and Federal Reserve Board Chairman Ben Bernanke regularly talks up his concern for small businesses and keep tabs on small business funding.").

270. *See* Lewis, *supra* note 269.

271. *See* Colleen DeBaise, *Kickstarting Entrepreneurship with "Startup America,"*

that features small business as a driver of economic recovery in the United States.[272]

2. Perceived and Actual Disadvantages of Crowdfunding

Currently, "crowdfunding is in a very early and noticeably unsettled state."[273] Comprised of a rapidly changing set of Internet business models, crowdfunding may be less transparent and more intangible to investors and regulators. Promoters of crowdfunding interests often are anonymous individuals and unknown entities. Moreover, in its prevalent current form as a small business start-up financing method, crowdfunding shares many of the overall negative attributes of small business and start-up capital formation.[274]

There are many traps for the unwary in this relatively new, rapidly developing, faceless transactional environment. Small businesses, especially start-ups, fail at a relatively high rate, and investors are likely to lose all of their investment.[275] In fact, it may be easier for investors who

---

WSJ.COM (Feb. 1, 2011, 11:40 AM ET), http://blogs.wsj.com/in-charge/2011/02/01/kick starting-entrepreneurship-with-startup-america/.

272. *See* Angus Loten, *'Startup America' Embraces Crowd-funding*, WSJ.COM (Apr. 22, 2011, 1:34 PM ET), http://blogs.wsj.com/in-charge/2011/04/22/'startup-america'-embraces-crowd-funding/; *see also* Lewis, *supra* note 269.

273. LAWTON & MAROM, *supra* note 4, at 71.

274. *See, e.g.*, Jill E. Fisch, *Can Internet Offerings Bridge The Small Business Capital Barrier?*, 2 J. SMALL & EMERGING BUS. L. 57, 58–64 (1998) (summarizing many of these attributes).

275. LAWTON & MAROM, *supra* note 4, at 180–81 ("The risk-reward curve in the startup world is quite well established. . . . At the risk of sounding too general, a lot of time most of the investment is lost . . . ."); *id.* at 58 ("Companies with small capitalizations present disproportionate risks of . . . business failure"); Brian Headd, et al., *What Matters More: Business Exit Rates or Business Survival Rates?*, 3, http://www.ces.census.gov/docs/bds/Exit%20Rates%20or%20Survival%20Rates.pdf ("The one-year survival rates for establishments born to firms started in 2004 was 76.4 percent and the five-year survival rates for establishments born to firms started in 2000 was 50.7 percent." (footnote omitted)). The U.S. Small Business Administration cites to the following findings about the survival of new firms, based on the same data used by Headd, *supra*:

> Seven out of 10 new employer firms last at least 2 years, half at least 5 years, a third at least 10 years, and a quarter stay in business 15 years or more. Census data report that 69 percent of new employer establishments born to new firms in 2000 survived at least 2 years, and 51 percent survived 5 or more years. Survival rights were similar across states and major industries. Bureau of Labor Statistics data on establishment age show that 49 percent of establishments survive 5 years or more; 34 percent survive 10 years or more; and 26 percent survive 15 years or more.

U.S. Small Bus. Admin., *Frequently Asked Questions: Advocacy: the Voice of Small Business in* Government, SBA.GOV, http://www.sba.gov/sites/default/files/files/sbfaq.pdf (last

lack corporate finance expertise or knowledge of relevant industries to lose their savings through online investments.[276] The SEC provides specific guidance to retail investors engaging in online trading:

> Online trading is quick and easy, but online investing takes time.
>
> With the click of a mouse, you can buy and sell stocks . . . . Although online trading saves investors time and money, it does not take the homework out of making investment decisions. You may be able to make a fast trade, but making wise investment decisions takes time. Before you trade, know why you are buying or selling, and the risk of your investment.[277]

Moreover, the Internet may over-inform and, as a result, obfuscate or bury important information in connection with securities offerings.[278] The Internet's capacity for encouraging suboptimal decision-making and the perceived higher probability of investor losses are real concerns. However, the Securities Act is not designed and does not exist to protect all investors from losing their money in all circumstances; the Securities Act is not an insurance policy against investor losses.[279] An insurance policy of that kind is neither realistically possible nor universally desirable.[280]

---

visited Dec. 26, 2011) (answering the question: "What is the survival rate for new firms?").

276. *See* LAWTON & MAROM, *supra* note 4, at 180 (noting that people commonly are concerned about "unsophisticated investors losing all of their money" in crowdfunding interests); Stephen J. Choi, *Gatekeepers and The Internet: Rethinking The Regulation of Small Business Capital Formation*, 2 J. SMALL & EMERGING BUS. L. 27, 37–38 (1998) ("With the increase in the number of active investors on the Internet comes a corresponding increase in potentially unsophisticated investors. . . . Because fraudulent issuers may sell securities with greater ease over the Internet, these investors are at risk." (footnote omitted)); Fisch, *supra* note 274, at 58 ("[R]isks may be magnified by Internet-based securities transactions. The low cost and wide distribution of Internet offerings makes the Internet an easy vehicle for fraudulent securities transactions." (footnote omitted)).

277. U.S. Sec. & Exch. Comm'n, *Online Investing*, INVESTOR.GOV, http://www.investor.gov/researching-managing-investments/investing-your-own/online-investing (last visited Dec. 26, 2011).

278. *See* Troy A. Paredes, *Blinded by the Light: Information Overload and its Consequences for Securities Regulation*, 81 WASH. U. L. Q. 417, 419 (2003) ("Studies show that at some point, people become overloaded with information and make worse decisions than if less information were made available to them." (footnote omitted)).

279. *See, e.g.*, *In re* Williams Sec. Litig.—WCG Subclass, 558 F.3d 1130, 1137 (10th Cir. 2009) ("The securities laws are not meant to 'provide investors with broad insurance against market losses, but to protect them against those economic losses that misrepresentations actually cause.'" (citation omitted)); Margaret V. Sachs, *Materiality and Social Change: The Case for Replacing "the Reasonable Investor" with "the Least Sophisticated Investor" in Inefficient Markets*, 81 TUL. L. REV. 473, 485 (2006) (noting that the materiality standard applicable to liability actions under, among other provisions, Section 17(a) of the Securities Act bases investor protection on the reasonable investor—"someone

A more specific concern is the capacity for fraud in crowdfunding.[281] Fraud protection is a focus of the Securities Act, and the Internet is a common vehicle for securities fraud.[282] Additional regulation is unlikely to significantly change this. As a result, the SEC has directed efforts toward investor education. For example, the SEC has a web page devoted to Internet fraud.[283] Also, small businesses may be disproportionately involved

---

who grasps market fundamentals").

280. The law's protection of people who unwisely part with money varies depending on context, and is based on a balancing of relevant policy interests. For example, crowdfunding advocates and analysts note the lack of parallel regulation of crowdfunding, gambling, and charitable solicitations, each of which may result in the loss of some or all of the funds conveyed by the people providing money. *See* LAWTON & MAROM, *supra* note 4, at 179–80 (noting that people can gamble away $5,000 without question, but cannot similarly invest $5,000 because of existing "American SEC regulation"); Angus Lotan, *Crowd-Fund Sites Eye Boom*, WSJ.COM (May 12, 2011), http://online.wsj.com/article/SB10001424052748703806304576245360782219274.html ("Until now, U.S. regulations permitted these [crowdfunding] sites only to facilitate donations—not purchases of equity stakes."); Gus G. Sentementes, *Crowdfunding Allows Everyone to Be an Arts Patron*, BALT. SUN (June 13, 2011), http://articles.baltimoresun.com/2011-06-13/business/bs-bz-crowdfunding-websites-20110613_1_arts-organizations-crowdfunding-arts-spending (describing charitable fundraising through crowdfunding and noting that "[f]ederal regulations prohibit fundraisers from using the websites to sell shares in projects to entice investors looking for a financial return").

281. LAWTON & MAROM, *supra* note 4, at 180 (indicating concern about crowdfunding's "potentials for fraudulent fund raising activities"). *See generally* Stephen Choi, *Regulating Investors Not Issuers: A Market-Based Proposal*, 88 CALIF. L. REV. 279, 308 (2000) ("Allowing truly unsophisticated investors to purchase securities of small speculative businesses, however, may lead to both mistake and fraud.").

282. Gregory C. Yadley, *General Solicitation: Looking for Funds in all the Wrong Places*, 70 FLA. B. J. 80, 81 (1996) (describing concern about securities fraud in "cyberspace"); Jake van der Laan et al., *Identifying Internet Mediated Securities Fraud: Trends and Technology*, WEBSCIENCE.ORG, 2 (Apr. 2010), http://journal.webscience.org/367/2/websci10_submission_71.pdf ("Over the last number of years North American securities regulatory agencies have noted a material increase in the number of securities fraud cases mediated through the internet." (footnotes omitted)). This issue is not new. Thirteen years ago, a respected securities law scholar made the following observations, which continue to be true today:

> Internet offerings present the risk of fraud. The media have publicized the popularity of the Internet as a tool for fraudulent transactions generally, and although Internet offerings are in their infancy, dishonest promoters have been quick to capitalize on the Internet's potential for cheating investors. The SEC has already identified and prosecuted promoters in connection with a variety of fraudulent Internet offerings, including pyramid schemes, false promises and sales of nonexistent securities.

Fisch, *supra* note 274, at 80 (footnotes omitted).

283. U.S. Sec. & Exch. Comm'n, *Internet Fraud*, SEC.GOV, http://www.sec.gov/

in securities fraud.[284] The SEC has recognized and reacted to the capacity for fraud in the market for securities of small businesses.[285] However, this enhanced potential for small business fraud does not mean we should ban, thwart, or unduly constrain securities offerings by small business issuers, even if those offerings occur over the Internet:

> Small companies have been responsible for a large proportion of the instances of investor fraud. By allowing small companies to make Internet offerings will we be giving the green light to the scam artists? No doubt more will try. Better investor education and stronger enforcement efforts should make the increase in fraud bearable, however. Moreover, the increase in fraud will be offset by the increase in legitimate business activity stimulated by the reduced costs of raising capital for many of our most innovative and productive companies.[286]

Given the prospect of investor losses resulting from small business failure and fraud and in spite of the potential capital formation efficiencies and other benefits it may create for small business, crowdfunding may, if unregulated or under-regulated, foster a lack of trust in the securities markets (or at least the crowdfunding component of those markets). Investor losses and fraud, as well as inconsistent business practices, may contribute to perceptions that the crowdfunding market is dishonest or corrupt. Any perception of market unfairness or distrust may have serious effects on investor confidence and investment behavior:

> In order for the securities markets to work, it is critical to maintain investor trust in the integrity of the market because this trust is the foundation on which the markets are built. Without a broad-based investor

---

investor/pubs/cyberfraud.htm (last modified Feb. 1, 2011).

284. *See* Choi, *supra* note 276, at 29; Richard J. Pierce, Jr., *Small Is Not Beautiful: The Case Against Special Regulatory Treatment of Small Firms*, 50 ADMIN. L. REV. 537 (1998); Fisch, *supra* note 274, at 58; David B. Guenther, Note, *The Limited Public Offer in German and U.S. Securities Law: A Comparative Analysis of Prospectus Act Section 2(2) and Rule 505 of Regulation D*, 20 MICH. J. INT'L L. 871, 908 (1999).

285. *See* Bradford, *supra* note 145, at 30.

286. Dale A. Oesterle, *The High Cost Of IPOs Depresses Venture Capital in the United States*, 1 ENTREPREN. BUS. L.J. 369, 379 (2006) (footnote omitted); *see also* Bradford, *supra* note 145, at 30 ("There may be more fraud in small offerings, or at least proportionately more, but that does not refute the basic argument: as long as many small offerings are legitimate, there is some offering amount below which registration, or any conditional exemption, is inefficient." (footnote omitted)); *id.* at 34 ("The SEC's concern with microcap fraud is laudable, but the Commission should not penalize all small business issuers for the misdeeds of a few. The answer to fraud lies in aggressive use of Rule 10b-5 and other antifraud rules, not in a prophylactic bar that ensnares even the smallest, least sophisticated businesses.").

> perception of legitimacy, people will not invest in the market, but put their money elsewhere, in "gold or real estate, or under their mattresses."[287]

The relative ease with which an unsophisticated investor may lose money in investments with small business issuers, the high rate of securities fraud in the small business context, and the anonymity of the Internet may give us pause about extending exemptive relief to crowdfunded offerings. However, crowdfunding has the capacity to fuel small business growth and satisfy the demand for a securities market that serves the everyman. Moreover, crowdfunding is a social and economic force not to be ignored: "Just how encompassing crowdfunding is, speaks to the enormity of its potential for economic and social impact. In the same way that social networking changed how we allocate time, crowdfunding will change how we allocate capital."[288] Accordingly, we believe that crowdfunding should be encouraged and crowdfunded offerings should be exempt from registration under the Securities Act; however, because it comes with both positive and negative consequences, a crowdfunding registration exemption should be cautiously pursued and appropriately tailored to accentuate the positive and minimize the negative.

B. The Contours of a Possible Crowdfunding Exemption

In an effort to take away unnecessary legal and regulatory barriers to crowdfunding while, at the same time, maintaining investor and market protections, we begin by establishing foundational principles and considering appropriate rulemaking options and processes. With these principles, options, and processes in mind, we then outline the possible parameters of a crowdfunding exemption from registration. The remainder of this Part sets forth those principles, options, processes, and parameters as a basis for further dialogue and action.

1. Foundational Principles

Although the following overlapping principles may seem simple and obvious, we consider each to be of importance in fashioning appropriate changes to crowdfunding regulation:

- Limit investor risk;
- Optimize fraud protection;
- Enhance informational transparency;
- Foster standardization of disclosures and enforcement;
- Constrain regulatory costs; and

---

287. Ripken, *supra* note 253, at 194 (footnote omitted).

288. LAWTON & MAROM, *supra* note 4, at 1.

- Minimize costs to issuers and investors.[289]

The first three of these principles derive directly from the policies underlying the Securities Act described in Part III.B. As corollaries of the Securities Act's investor protection and market integrity maintenance objectives, these principles are central values for any variance in the current securities regulation scheme. Among other things, the mandatory disclosure rules exist to promote transparency for the protection of investors, prevent fraud, and promote the perception (if not the reality) of fair and honest markets.[290] Moreover, the three initial principles go to the heart of the matters described in Part IV.A.2 as regulatory concerns applicable to crowdfunding: specifically, unease about the relationship between Internet offerings and investor losses and higher probabilities of fraud in small business capital formation.[291] Transparency—meaning not necessarily more disclosure, but more targeted, simple, easy-to-access disclosure—should support more effective transmission of information to the potentially inexperienced or less experienced Internet investors that are among those attracted to crowdfunding.[292] In addition, transparency should help limit cases of fraud to crowdfunded businesses that are affirmative bad actors whose conduct is not likely to be deterred by ex ante regulation and must be punished by ex post enforcement.

The fourth value, standardization, is closely related. Standardization of disclosures through mandatory disclosure requirements and enforcement is an efficient way to accomplish the first three principles, but especially relates to and operates hand-in-hand with transparency. In addition, disclosure and enforcement standardization is a potential source of economic efficiencies that may help constrain costs incurred by market participants (the fifth and sixth principles in our list of foundational principles) while also protecting investors:

---

289. We note that Woodie Neiss, a member of the Small Business and Entrepreneurship Council Advisory Committee, suggested four similar principles as a basis for crowdfunding regulation in a December 2010 letter appended to the Final Report of the 2010 Forum on Small Business Capital Formation. *See* SEC, 2010 ANNUAL SEC GOVERNMENT-BUSINESS FORUM ON SMALL BUSINESS CAPITAL FORMATION: FINAL REPORT (June 2011), http://www.sec.gov/info/smallbus/gbfor29.pdf.

290. *See* Judge Stanley Sporkin, *The Worldwide Banning of Schmiergeld: A Look at the Foreign Corrupt Practices Act on its Twentieth Birthday*, 18 NW. J. INT'L L. & BUS. 269, 272 (1998) ("The securities laws have long been a model for appropriate government regulation. They are largely statutes that mandate transparency. Full and fair disclosure is the general concept underpinning these laws."); Kevin Werbach, *Sensors and Sensibilities*, 28 CARDOZO L. REV. 2321, 2345 (2007) ("A great deal of securities regulation revolves around mandatory disclosure, in order to promote transparency of financial markets.").

291. *See supra* Part IV.A.2.

292. *See* Lotan, *supra* note 280 (noting that "crowd-funding sites advertise start-ups to a broad audience, and could easily attract people who shouldn't be involved in speculative offerings").

> Absent standardization of disclosure requirements, such as those provided by the SEC, there will remain "grave uncertainty about outcomes because such matters as intent and negligence need to be sorted out in court." By contrast, the securities regulatory system is standardized, which makes disclosure more efficient.
>
> The standardization of disclosure and enforcement rules is a central benefit for securities regulation and a key reason why private contracting cannot as efficiently protect the investor.[293]

Although a number of respected law and economics scholars advocate voluntary disclosure rather than mandatory disclosure,[294] because standardized disclosures exist for issuers outside the crowdfunding context (e.g., through registration requirements and Regulations D and A[295]), we also include standardization as a foundational regulatory principle here. We want to be clear, however, that fostering standardization does not necessarily require homogenization or the imposition of weighty line-item disclosure rules.

Finally, our listed foundational principles recognize that it is important, in our attempt to both promote crowdfunding and protect investors and markets, not to impose on regulators or market participants a level of cost that is perceived to be so prohibitively high that the proposed regulatory solution is not feasible for regulators or represents a disincentive for issuers or investors. The cost of complying with regulatory change is an important consideration. Certain costs, such as the resources devoted to completing necessary filings and disclosures, are obvious.[296] However, the assessment of those costs in a crowdfunding context may be difficult, especially in light of crowdfunding's varied forms and relatively short track record. Moreover, current regulatory cost-benefit analyses for federal agency rulemaking may not be as broadly applicable as they should be or, when applicable, may be unavailing for other reasons.[297] Accordingly, additional means of assessing

---

293. Frank B. Cross & Robert A. Prentice, *The Economic Value of Securities Regulation*, 28 CARDOZO L. REV. 333, 356–57 (2006) (footnotes omitted).

294. *See, e.g.*, Stephen M. Bainbridge, *Mandatory Disclosure: A Behavioral Analysis*, 68 U. CIN. L. REV. 1023, 1024–25 (2000) (setting forth both neoclassical economic and behavioral aspects of the mandatory disclosure debate).

295. *See supra* notes 143, 194, 213 and accompanying text.

296. *See supra* note 124 and accompanying text.

297. *See generally* CASS R. SUNSTEIN, THE COST-BENEFIT STATE: THE FUTURE OF REGULATORY PROTECTION (2002) (defending the use of cost-benefit analysis in regulatory rule-making, but criticizing its misuse and suggesting alternative approaches); Susan Rose-Ackerman, *Putting Cost-Benefit Analysis in Its Place: Rethinking Regulatory Review*, 65 U. MIAMI L. REV. 335 (2011) (critiquing and suggesting revisions to the existing means of engaging cost-benefit analysis); Arden Rowell, *The Cost of Time: Haphazard Discounting and the Undervaluation of Regulatory Benefits*, 85 NOTRE DAME L. REV. 1505 (2010) (criticizing regulatory cost-benefit analysis on the basis of, among other things, time indeterminacy).

the costs and benefits of a crowdfunding exemption should be identified, explored, and potentially used to supplement existing methods.[298]

These additional cost-benefit analyses should take account of the costs associated with the fact of a change in legal rule, in addition to costs associated with the specific substantive attributes of the proposed new rule.[299] These legal transition costs frequently go unnoticed and unaccounted for in assessing the overall costs associated with a new legal rule:

> What has been overlooked is the friction inherent in change itself. Whatever one's normative perspective, a legal system will incur costs simply in adjusting to the existence of a new legal norm. These will arise, for instance, from the need to learn about the content of new law, as well as from an increased risk of uncertainty about its meaning and effect. Changes in legal directives likewise will compel intraparty adjustments and have subtle effects on interparty relationships forged around the old legal order. Indeed, transition costs reflect a systemic phenomenon. Although in differing degrees, they will arise from legal change in all fields, with all lawmaking structures (whether statutory, administrative, or judicial), and for all types of reform (regulatory, deregulatory, and so on).[300]

These costs are inevitable in any legal rule change, but can be greater in some rule changes than others.[301] In the case of a possible new crowdfunding registration exemption, legal transition costs may be borne by regulators, issuers, investors, and others and may include:

- those associated with lawyers, judges, and other legal professionals, principals of crowdfunding websites and crowdfunded ventures, and others learning about the new exemption;
- those borne by people with knowledge of and experience with the old rule (trading off their own knowledge of SEC registration exemptions for an untested new exemption) and by people who struggle with interpretive questions about and possible gaps in the new registration exemption regime;
- lost opportunities resulting from a lack of clarity or precision in the expression of the new exemption;
- the expense of creating new forms and business practices to comply with the new exemption;
- those incurred by legal actors who misunderstand the new exemption and act in reliance on that misunderstanding in creating faulty new

298. *See, e.g.*, Bradford, *supra* note 145, at 30–33 (modeling the costs and benefits of registration or partial regulation, and exemption).

299. *See generally* Michael P. Van Alstine, *The Costs of Legal Change*, 49 UCLA L. REV. 789 (2002) (cataloguing and categorizing legal transition costs).

300. *Id.* at 793.

301. *Id.* at 816.

forms or business practices, in giving erroneous advice on the application of the new exemption, or in conducting their crowdfunding businesses; and

- those suffered by the SEC in administering and applying the new exemption.[302]

The SEC can reduce these types of costs by constructing any new crowdfunding exemption in a way that minimizes variation from the existing registration exemption scheme and otherwise decreases uncertainties and misunderstandings. Use of the existing exemptive framework and prevailing industry norms and best practices, for example, may best serve this purpose.

2. Rule-Making Options and Processes

Having established the basic values that should underlie any proposed changes to crowdfunding regulation, we next consider the options and process for creating a new registration exemption for crowdfunding consistent with those values. Several possible paths are immediately apparent. Because we assume that profit-sharing crowdfunding interests are securities under the Securities Act and that regulatory change is required to serve our interest in facilitating crowdfunded offerings,[303] we excluded maintenance of the status quo (in which no registration exemption is available) as a possible regulatory option.

The first possible regulatory option is to treat crowdfunding as a completely new and distinct business model and regulate it separately from other capital formation activities. The most simple and comprehensive way to accomplish this would be for the SEC to provide, by regulation or interpretive guidance, that crowdfunding interests are not securities for purposes of the Securities Act and the Exchange Act.[304] Absent further regulation or guidance, this approach would render the registration requirement in Section 5 of the Securities Act[305] inapplicable to crowdfunded offerings. This method of facilitating crowdfunding would minimize costs to issuers, investors, and the SEC because it leaves any regulation of crowdfunding to the market. This course of action may, however, present more perceived or actual risk to funders and create moral hazard by decreasing protections against fraud (since antifraud statutes and

---

302. *Id.* at 850–52 (describing and illustrating the various forms of legal transition costs on which this list is based).

303. *See supra* Part IV.A.

304. *See supra* Part II. Although the definition of a security differs under the Securities Act and the Exchange Act, we believe that treating the two definitions similarly in this context makes sense because the current definitions are substantially the same, and the regulatory environments are interwoven. *See supra* note 141.

305. *See supra* notes 133, 135 and accompanying text.

rules under federal securities regulation would be inapplicable).[306] Moreover, while the effects of market self-regulation on informational transparency and the standardization of disclosure and enforcement may be theorized in this context, they are unclear.

One way to address the uncertainties created by deregulating the offer and sale of crowdfunding interests under the federal securities laws is by regulating those transactions under another one or more existing areas of law (e.g., through gambling regulation or the regulation of charitable donations[307]) or by regulating them under a new scheme of regulation created especially for crowdfunding. If this option were pursued, existing regulatory frameworks would likely have to be significantly modified to comply with our articulated principles of regulatory change, and creation of a new regulatory scheme consistent with these principles would consume significant time and resources. Both of these options—modifying and using a pre-existing regulatory scheme and creating a new regulatory scheme—increase regulatory costs and costs to issuers and investors. These options also would have to be designed and implemented to optimize fraud protection, enhance informational transparency, and foster disclosure and enforcement standardization.

The more promising regulatory option is to clarify that some or all crowdfunding interests are securities under federal law and to work within the existing securities regulation scheme to fashion a registration exemption that is consistent with the foundational principles articulated in Part IV.B.1. First, by working in the existing regulatory framework, we can constrain both regulatory costs and costs to issuers and investors. Second, through customized provisions, the architects of the exemption can work to limit investor risk, optimize fraud protection, enhance informational transparency, and foster disclosure and enforcement standardization. This seems like the best approach.

Of course, this path is not as easy as it sounds. As we noted earlier, crowdfunding incorporates a variety of different business models,[308] and it

306. Prosecutors, federal agencies, and aggrieved investors still may have paths to enforcement under state securities (or "blue sky") laws and regulations as well as federal and state fraud law outside the securities fraud context. *See* David M. Cielusniak, Note, *You Cannot Fight What You Cannot See: Securities Regulation on the Internet*, 22 FORDHAM INT'L L.J. 612, 634–35 (1998). These other fraud actions may or may not have desirable remedies or advantageous claim elements for public enforcement authorities or a particular aggrieved private investor, however. By cutting down the number of potential enforcement avenues, perceptions of both investor protection and market integrity may suffer.

307. For a brief "take" on the gambling and securities analysis of crowdfunding on the music industry, *see* Tim Kappel, Comment, Ex Ante *Crowdfunding and the Recording Industry: A Model for the U.S.?*, 29 LOY. L.A. ENT. L. REV. 375, 382–83 (2009); *see also* Hazen, *supra* note 7, at 1, 20 n.29.

308. *See supra* note 4 and accompanying text; *see also* LAWTON & MAROM, *supra* note 4, at 201 ("It's worth noting that there are many categories of crowdfunding, and it would

is still a new phenomenon with uncertain advantages and disadvantages.[309] Experience in crowdfunding is dispersed and disparate, and no individual or group of individuals in the SEC is likely to have sufficient expertise to piece together an appropriate exemptive solution:

> [R]egulators and policy makers are just waking up to crowdfunding. There have been a limited number of regulatory shut-down events in the recent years, but by and large regulators need to play catch-up. . . . [T]he velocity of this space is accelerating, and even for someone who spends full-time looking into is [sic], it's impossible to keep track of it all. How people who have other roles and responsibilities . . . could properly regulate crowdfunding is hard to imagine.[310]

Accordingly, it seems prudent to engage those involved in crowdfunding in the regulatory discussions in a meaningful way before Congress passes legislation that legalizes crowdfunded offerings of securities or the SEC publishes a rule proposal as part of the notice-and-comment process required under the U.S. Administrative Procedure Act ("APA").[311] SEC Chairman Mary Schapiro embraced this idea in connection with the Dodd-Frank Wall Street Reform and Consumer Protection Act and is again embracing it in the context of reforms to small business capital formation (of which crowdfunding regulation may become a part).[312] We endorse this approach.

---

not be appropriate to apply the same disciplines across them all.").

309. *See supra* Part IV.A.

310. *See* LAWTON & MAROM, *supra* note 4, at 198. Although two industry researchers suggest self-regulation of crowdfunding, *see id.* at 198–202, we are not persuaded that this is a wise course of action given the foundational principles we articulate *supra* Part IV.B.1. Our discussion *supra* in Part IV.B.2 is instructive in this regard. Even these two researchers offer that regulators should be involved in the process. *Id.* at 198–99. The question is who controls, monitors, and enforces the regulation.

311. *See* 5 U.S.C. § 553 (2006). Under the APA, "[a]fter notice required by this section, the agency shall give interested persons an opportunity to participate in the rule making through submission of written data, views, or arguments with or without opportunity for oral presentation." *Id.* at § 553(c).

312. *See* Schapiro Testimony, *supra* note 11 (noting that, in conducting a review of the regulation of small business capital formation, the SEC "will gather data and seek input from many sources, including small businesses, investor groups, the public-at-large, and a new Advisory Committee on Small and Emerging Companies that the Commission is in the process of forming, so that we consider a variety of viewpoints."); Mary L. Schapiro, Chairman, SEC, Remarks Before the Society of American Business Editors and Writers (Apr. 8, 2011), http://www.sec.gov/news/speech/2011/spch040811mls.htm (reflecting on regulatory efforts undertaken by the SEC under Dodd-Frank and noting that "[q]uality rules can't evolve in a Washington bubble. We understand the impact our actions can have on the financial markets, on companies large and small, and on individual lives. This understanding drives us to hear a wide range of opinions, and consider every view as we move forward to carry our mission . . . .").

The involvement of crowdfunded businesses, promoters, and investors in the creation of a crowdfunding exemption may have more than a substantive advantage, however. There may be positive cognitive effects to employing industry participants in the regulation process. There is a proven value, in the form of buy-in, when members of a group are asked to comply with direction by members of their in-group.[313] In addition, introducing non-SEC personnel into the regulatory decision-making process may mitigate the effects of bounded search, bounded rationality, groupthink, and other operative behavioral biases.[314] A regulatory process that achieves buy-in to mutually acceptable rules and engages fresh, proactive decision-making by incorporating regulated businesses into the rulemaking process early may better serve the foundational principles of limiting investor risk, optimizing fraud protection, enhancing informational transparency, and standardizing disclosure and enforcement. Positive effects on costs also may result.

While potentially advantageous for all these reasons, integrating crowdfunding representatives with regulatory authorities in the rulemaking process must be done carefully to avoid both inefficiencies in process (which may result from the involvement of disparate industry participants in the regulatory process) and the perceived or actual co-opting of regulators by industry—so-called regulatory capture.[315] We have considered the use of a wiki or other form of crowdsourced regulatory initiative as a potential solution to these problems and as a possible cost-saving device, but we have concluded that this vehicle, taken alone, likely would not provide the kind of detailed input, debate, and discussion that group working meetings featuring simultaneous (or perhaps, synchronous) interaction conducted in person or electronically would provide. Further, we have determined that it may be prudent to introduce the collaboratively developed regulatory response as a pilot program for further study after a period of years. While this would increase regulatory costs, it seems unlikely, given the fast rate at which crowdfunding is developing, that an

313. *See* LYNN STOUT, CULTIVATING CONSCIENCE: HOW GOOD LAWS MAKE GOOD PEOPLE 145–46 (2010).

314. *See* Stephen J. Choi & A.C. Pritchard, *Behavioral Economics and the SEC*, 56 STAN. L. REV. 72–73 (2003).

315. *See* Choi, *supra* note 276, at 40 ("Regulators may also be subject to possibilities of regulatory capture from the very groups that the regulators seek to regulate."); Saule T. Omarova, *Wall Street as Community of Fate: Toward Financial Industry Self-Regulation*, 159 U. PA. L. REV. 411, 463 (2011) ("Regulatory agencies in charge of the financial services sector often display strong signs of industry capture and increasingly engage in nontransparent and highly informal rulemaking that falls outside public scrutiny and tends to favor the industry."). *See generally* Wendy E. Wagner, *Administrative Law, Filter Failure, and Information Capture*, 59 DUKE L.J. 1321 (2010) (describing and illustrating the role of information capture in agency rulemaking and enforcement).

initial regulatory response—even with industry input—would be able to fully anticipate the actions and reactions of industry participants.[316]

3. Substantive Elements of a Proposed Exemption

In a July 1, 2010 petition to the SEC, the Sustainable Economies Law Center ("SELC") advocates adopting a rule that would provide a registration exemption for offerings up to $100,000, with a maximum of $100 per investor.[317] With crowdfunded ventures as the target of this exemption, the petition specifies that the entity seeking funding, which the SELC refers to as the "offeror," must be an individual residing in the United States and may only have one offering open at a time.[318] In its petition, the SELC asserts that such an exemption will promote entrepreneurship and allow small businesses to raise equity rather than debt.[319] According to the SELC, these are both desired consequences, as they will stimulate economic recovery.[320] The SELC petition concludes by noting that Section 3(b) and Section 4(2) are possible sections under which the rule could be promulgated, but acknowledges that if the rule is created under Section 4(2), no general solicitation or advertising will be permitted.[321] SEC Chairman Mary Schapiro references the petition in her

---

316. We note that two prominent scholars have made similar suggestions for comparable and related reasons in the weeks immediately preceding the final editorial changes to this article. *See* Examining Investor Risks in Capital Raising: Hearing before the Subcommittee on Securities, Insurance, and Investment of the Sen. Comm. on Banking, Housing & Urban Affairs, 112th Cong. (2011) (testimony of Professor John C. Coates IV, John F. Cogan, Jr. Professor of Law and Economics, Harvard Law School), at 5, *available at* http://banking.senate.gov/public/index.cfm?FuseAction=Files.View&FileStore_id=1d24b42e-3ef8-4653-bfe8-9c476740fafa (suggesting a "sunset provision" for capital formation proposals in various pending bills, including the Senate crowdfunding proposals, "such that the proposals would by their terms last for no more than two or three years"); Roberta Romano, Regulating in the Dark (Dec. 18, 2011) (unpublished manuscript), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1974148 (arguing for, among other things, congressional and regulatory re-examinaton and re-evaluation of decisions to adopt proposals embodied in crisis legislation). Specifically, Professor Romano contends that Congress and regulators should provide "procedural mechanisms that require automatic subsequent review and reconsideration of those decisions, along with regulatory exemptive or waiver powers that create flexibility in implementation and encourage, where possible, small scale, discrete experimentation to better inform and calibrate the regulatory apparatus." *Id.* at 3–4.

317. Petition from Jenny Kassan, Co-Director, Sustainable Econs. Law Ctr. to Elizabeth M. Murphy, Sec'y Sec. Exch. Comm'n (July 1, 2010), *available at* http://www.sec.gov/rules/petitions/2010/petn4-605.pdf.

318. *Id.*

319. *Id.*

320. *Id.*

321. *Id.*

April 2011 letter to The Honorable Darrell E. Issa, Chairman of the Committee on Oversight and Government Reform of the U.S. House of Representatives.[322]

The SELC petition caught the attention of both crowdfunding advocates and the SEC,[323] and it has served a valuable role as a catalyst of efforts for change.[324] Although the petition includes certain key tenets of a possible crowdfunding exemption, it represents a unilateral, incomplete response to the issues crowdfunding raises under the Securities Act. In particular, while the inclusion of aggregate and per-investor caps has merit as a risk-reduction device, we are believe that more can be done in this regard at a relatively low cost to the SEC, issuers, and investors. Moreover, the SELC proposal does not address fraud protection, informational transparency, or disclosure and enforcement standardization. As discussed in Part IV.B.1, we believe that these principles are central to an appropriate, successful regulatory response. For example, we find it unacceptable for a crowdfunding regulatory exemption to leave those who invest a small dollar value in a venture to fend for themselves,[325] even though we allow individuals to bear the loss of the same amount of funds in gambling transactions or as financially ill-advised charitable donations or consumer purchases.[326] Generally, that approach would neither promote nor support market integrity, even if investor protection is deemed unnecessary for those advancing a small amount of capital. Finally, the SELC petition does not address a means for engaging crowdfunding participants in the process of constructing the exemptive proposal[327] (although crowdfunding issuers, investors, and advocates would have the opportunity to participate in the overall regulatory effort through the notice-and-comment process under the APA, as earlier noted[328]).

We also are aware of three other proposals (one being an article cautioning against the establishment of a crowdfunding exemption without

---

322. *See* Issa Letter, *supra* note 9.

323. *Id.*

324. *See* Bradford, *supra* note 4, at 52–53 (citations omitted).

325. *See* Bradford, *Crowdfunding Blog*, *supra* note 9 ("Obviously, an offering isn't any safer just because a large number of people invest small amounts. And people who invest small amounts aren't necessarily sophisticated enough to protect themselves; in fact, smaller investors are probably less sophisticated on average."). We are mindful, however, that the more limited protections afforded to accredited investors under Regulation D rely on the ability of accredited investors to bear the potential loss of their entire investment. *See supra* notes 193, 198 and accompanying text; *see also* Bradford, *Crowdfunding Blog*, *supra* note 9 (noting that the "argument for an exemption if investors can afford to lose the money isn't as novel as it sounds. It essentially underlies Regulation D's designation of investors as 'accredited' if they meet specified income or net worth limits.").

326. *See supra* note 280.

327. *See supra* notes 311–15 and accompanying text.

328. *See supra* note 311 and accompanying text.

a mandatory disclosure component), each of which appears in an article authored by a fellow law scholar at or about the time this article is being published.[329] Other proposals also have been advanced, and likely will continue to be advanced, on both formal and informal bases.[330] Each proposal has attributes common to others, and each proponent makes important arguments. Although we believe there is value in reading and considering each of these proposals in the regulatory process, we do not support any of them as a precise template for congressional or SEC rulemaking.

As earlier indicated in this article, the U.S. Congress has begun to take action.[331] Both the bill passed in the U.S. House of Representatives in November 2011 and the Bill under consideration in the U.S. Senate at the time this article went to press call for a specific crowdfunding exemption. Yet, there are significant differences between the two bills. The House bill sets an aggregate offering limit of $1,000,000 ($2,000,000, if the issuer furnishes investors audited financial statements) and a per-investor cap of the lesser of $10,000 or 10% of the investor's annual income.[332] The Senate bill caps the maximum aggregate offering price at $1,000,000 and sets a per-investor limit of $1,000.[333] The Senate bill requires that securities be issued through an intermediary, but the House bill does not.[334] Both bills require certain disclosures, restrict resales of the subject crowdfunding interests, and incorporate "bad boy" disqualifiers.[335]

Each proposal and bill has merits and flaws, and no doubt each will develop further after this article goes to press. In our view, however, the collaborative process we suggest in Part IV.B.2 represents a constructive and optimal approach to determining the specific terms of a proposed registration exemption for crowdfunding interests that is consistent with the

329. *See* Bradford, *supra* note 4; Hazen, *supra* note 7; Pope, *supra* note 215.

330. *See* Bradford, *supra* note 4, at 3, 51–56.

331. *See supra* notes 13–15 and accompanying text. The North American Securities Administrator Association also has released a proposal through Jack Herstein's testimony in the December 1 hearings held on S. 1791. *See Spurring Job Growth Through Capital Formation While Protecting Investors: Hearing on S. 1791 Before the S. Comm. on Banking, Housing, and Urban Affairs*, 112th Cong. (Dec. 1, 2011) (written testimony of Jack E. Herstein, President of the North American Securities Administrators Association, Inc. and Assistant Director of the Nebraska Department of Banking & Finance, Bureau of Securities), *available at* http://banking.senate.gov/public/index.cfm?FuseAction=Files.View&FileStore_id=255a1e89-30b9-4036-9560-b4a0db5def80.

332. H.R. 2930, 112th Cong. § 2(a) (2011) [hereinafter H.R. 2930], *available at* http://thomas.loc.gov/home/thomas.php (actual bill text may be accessed by searching for appropriate bill number).

333. S. 1791, 112th Cong. § 2(3) (2011) [hereinafter S. 1791], *available at* http://thomas.loc.gov/home/thomas.php (actual bill text may be accessed by searching for appropriate bill number).

334. H.R. 2930, *supra* note 332.

335. *Id.*; S. 1791, *supra* note 333.

foundational principles of limiting investor risk, optimizing fraud protection, enhancing informational transparency, fostering disclosure and enforcement standardization, constraining regulatory costs, and minimizing costs to issuers and investors—principles of regulation that we believe will allow the proper balance of governmental and market-based regulation.[336] We also suggest that those engaged in crafting an appropriate exemption consider proposing a new rule-based exemption using some or all of the following general substantive attributes:

- Limit and tailor the exemption to Internet offerings;
- Do not allow foreign issuers, investment companies, or public companies (i.e., issuers who do not have a class of securities registered under Section 12 of the Exchange Act[337]) to use the exemption;
- Permit general solicitation and general advertising;
- Limit the aggregate offering price for each crowdfunded venture to a specified dollar amount (e.g., $100,000, $250,000, or possibly even lower thresholds for some types of exempt offerings[338]) over a twelve-month period;[339]
- Limit the aggregate dollar value of crowdfunding interests that a single investor may purchase in a single crowdfunded venture (e.g., $100 or $250) in a single offering or over a specified period,[340] unless

---

336. In saying this, we acknowledge that others may have a different set of values, consistent with the policy underpinnings of the Securities Act, that they desire to promote in this regulatory process. Regardless, we would hope that the group crafting any exemption in this area would develop and articulate a set of principles to guide its activities.

337. 15 U.S.C. § 78l(g) (2006).

338. Regulation A formerly permitted offerings of up to $100,000 with limited required mandatory disclosures. *See* 17 C.F.R. § 230.257 (1991). Moreover, early guidance from the SEC's General Counsel noted that small aggregate offering size is a characteristic of a private placement exempt from registration under Section 4(2). *See* SEC General Counsel Letter, *supra* note 164. Also, we note that one academic proposal would limit the use of an exemption to "microstartups"—businesses "in which one or two creative people have an idea for a product or service that can be developed, launched, and marketed for a few thousand dollars." Pope, *supra* note 215, at 975.

339. We do not, however, suggest that the SEC consider whether twelve months is the appropriate aggregation period. *See infra* notes 346–48 and accompanying text.

340. Those crafting the regulatory exemption must be mindful of the relationship between the per-investor and per-offering caps, because crowdfunded ventures and crowdfunding websites will not want to trigger registration requirements under Section 12(g)(1) of the Exchange Act, 15 U.S.C. § 78l(g)(1) (2006), and Rule 12g-1, 17 C.F.R. § 240.12g-1 (2011). Together they require registration under the Exchange Act if an issuer has a class of equity securities held by at least 500 people and at least $10,000,000 in total assets. *See* Pope, *supra* note 215, at 996–97. Many business ventures that we envision using a crowdfunding registration exemption under the Securities Act would not have sufficient total assets to trigger registration under the Exchange Act. However, regulators must give attention to this issue.

the investor is an accredited investor or sophisticated (as those concepts are defined and used in Regulation D);
- Restrict resales of crowdfunding interests;
- Compel issuers (both crowdfunded ventures and their promoters, including crowdfunding website operators[341]) to file a brief issuer registration and a brief offering notice with the SEC; and
- Mandate that certain limited disclosures, including cautionary language, be included on the crowdfunding website in a specified manner and, in some cases, using specific text.

We expect that the precise combination of these attributes that will best effectuate desired foundational principles will be the subject of significant discussion and debate among participants in the regulatory process. Our approach encourages a balancing of issuer, investor, and regulatory interests in a manner similar to that involved in federal consumer protection regulation.[342] The overall analogy to consumer protection is too complex to explore in any depth here. Suffice it to say, however, that there are both commonalities and differences in selling securities and other products at similar price points over the Internet.[343] In determining the substantive

341. For this Part IV.B.3, we treat crowdfunding websites as co-issuers for most purposes. *See supra* Part III.A.3. By imposing regulation on crowdfunding websites as well as crowdfunded ventures, crowdfunding websites should be incentivized to engage in rigorous pre-screening of the crowdfunded ventures they host and promote. We note, however, that for our solution to work in this co-issuer context, a number of details would need to be addressed. For example, as we note earlier in this list of potential substantive attributes, any aggregate offering cap would have to apply to each crowdfunded venture, and not to each crowdfunding website (because there are crowdfunding websites that host and promote the securities of multiple crowdfunded ventures). Alternatives exist to treat the crowdfunding websites as brokers, investment advisors, exchanges, or other investor fiduciaries. *See supra* note 250 and accompanying text. The issue of how to treat promoters of small business offerings has been on the SEC's radar screen for quite some time. *See* Cohn & Yadley, *supra* note 144, at 61–63.

342. One commentator explained this balancing well in the context of the consumer protection regulation of ecommerce transactions:

> Government has an interest in ensuring that e-consumer confidence reinforces e-commerce as a viable commercial medium, benefiting both e-businesses and e-consumers. . . . Yet, government e-commerce action must be calculated and targeted, balance market and social policies in the process, take the Internet mechanism into consideration and not eliminate e-commerce's attraction—efficiency, low cost, easily accessible consumer base and the simultaneous nature of business transactions.

John R. Aguilar, *Over the Rainbow European and American Consumer Protection Policy and Remedy Conflicts on the Internet and a Possible Solution*, 4 INT'L J. COMM. L. & POL'Y 1, 10–11 (1999) (footnotes omitted).

343. Securities regulation in the crowdfunding context may be seen as a specific form of consumer protection, emphasizing investor protection policy over market protection and

attributes of an appropriate exemption for crowdfunding, we suggest that rule makers examine the interactions among those attributes in the context of the policies underlying federal securities regulation while, at the same time, keeping overall consumer protection principles in mind. The remainder of this Part IV.B.3 sets forth further thoughts that may be relevant to a consideration of those interactions.

Ultimately, we are not convinced of the need for Congress to act on a crowdfunding exemption under the Securities Act (although Congress will likely need to take parallel action on other securities regulation issues in order to effectuate the exemption in full[344]). Although a thorough comparative institutional analysis is beyond the scope of this article, we note that existing regulation offers ample opportunity for the SEC to act without a grant of additional congressional authority (and, presumably, at a lower aggregate cost) and that the SEC's overall competence and relative independence make it a desirable rule maker in this context.[345] Specifically, we contemplate that the SEC would use its exemptive authority under Section 3(b) of the Securities Act[346] to promulgate this exemption, which could be included in Regulation D as, e.g., Rule 504A, or in a new parallel regulation modeled after Regulation D (perhaps denominated Regulation CF). The concepts of integration[347] and aggregation[348] applicable to Rule 504 and 505 offerings under Regulation D also would be applicable to exempt crowdfunded offerings, although we recommend that the SEC consider shortening the periods for each, consistent with its rule-making authority under Section 3(b) and Section 28 of the Securities Act.[349] By working within the existing regulatory framework for Section 3(b)

---

promotion. *See* DEE PRIDGEN & RICHARD M. ALDERMAN, CONSUMER PROTECTION AND THE LAW: 2009–2010 EDITION § 1:1 (West 2009) (noting a common "philosophy that the government should play a role in assuring that consumers are not unfairly taken advantage of in the marketplace."). The commonalities between securities regulation in this context and consumer protection regulation in the crowdfunding context extend to, among other things, protections against fraud, deceit, misrepresentation, and deceptive and unfair practices. *See generally id.* at §§ 2:1, 3:1 (describing these regulatory areas under consumer protection law).

344. *See infra* notes 388–89 and accompanying text.

345. *See generally* Joan MacLeod Heminway, *Rock, Paper, Scissors: Choosing the Right Vehicle for Federal Corporate Governance Initiatives*, 10 FORDHAM J. CORP. & FIN. L. 225 (2005) (suggesting a framework for a comparative institutional analysis of federal corporate governance initiatives).

346. *See supra* note 184.

347. *See supra* note 187 and accompanying text.

348. *See supra* note 190 and accompanying text.

349. *See* Cohn & Yadley, *supra* note 144, at 47–54; Hazen, *supra* note 7, at 10, n.66. The SEC proposed shortening the integration window from six months to ninety days in 2007. *See* Revisions of Limited Offering Exemptions in Regulation D, Securities Act Release No. 33-8828, 72 Fed. Reg. 4516 (proposed Aug. 3, 2007), *available at* http://www.sec.gov/rules/proposed/2007/33-8828.pdf.

exemptions under Regulation D, regulatory costs should be less than if new regulations were created from whole cloth.

The substantive characteristics set forth above are designed to operate in the limited context of very small Internet-based offerings for relatively small U.S. issuers—a context in which crowdfunding has its perceived maximum net advantages under current circumstances and in which U.S. law applies.[350] Although the substance of our proposal may be faulted for representing a somewhat timid response, it is "much better than nothing."[351] We may be wrong, but given the relatively novel nature of crowdfunding and the existing state of the SEC, we believe that a conservative initial approach is warranted.[352] There may come a day on which it would be appropriate to extend a crowdfunding exemption to larger offerings in wider contexts; however, we are concerned that the multiplicity of crowdfunding models and flux in current crowdfunding platforms make it too difficult to fashion a wider exemption that adequately limits investor risk and optimizes fraud protection.[353] There also may be a future time at which it would be advisable to initiate an overhaul of all small business capital formation regulation. The current investor-protection focus of the SEC and its lack of adequate funding[354] make this an improbable current objective. Accordingly, as a more limited approach, we suggest an Internet small issuer exemption that limits the aggregate dollar value of offerings covered, and we recommend limiting the availability of the exemption to non-public U.S. issuers that are not investment companies[355] using open-access websites to enhance their base of prospective and actual funders.

---

350. *See supra* Part IV.A.1.

351. LAWTON & MAROM, *supra* note 4, at 188 (noting that "from a macro view, exemptions like this can be counter-productive, and in fact might hold back crowd-funding's potential").

352. Professor Thomas Lee Hazen takes an even more conservative approach than the one we suggest here. In general, Professor Hazen does not favor the creation of a new, broad exemption for crowdfunding. Hazen, *supra* note 7, at 17–22. However, he does offer that "[i]n the event that an additional exemption is warranted, it should be conditioned on mandated disclosures that would give investors the opportunity to evaluate the merits of the investment." *Id.* at 22. The type of disclosure Professor Hazen has in mind is disclosure akin to that provided in a Regulation A offering, a more weighty level of disclosure than we recommend here. *Id.* at 14–15, 22.

353. *See* LAWTON & MAROM, *supra* note 4, at 93–102 (describing actual and aspirational attributes of crowdfunding platforms, and noting that the characteristics of a potential future crowdfunding platform may make regulators "a lot more comfortable," and better support necessary fraud prevention algorithms).

354. *See generally* Heminway, *supra* note 345 (discussing the SEC's investor-protection mission and funding situation in the context of an evaluation of reform efforts at the SEC).

355. We note that neither investment companies nor public companies can use Rule 504 to avoid registration under the Securities Act and that Rule 505 is not available to investment companies. *See* 17 C.F.R. §§ 230.504(a), 505(a) (2011); *see also* Campbell, *supra* note 146,

Our focus on open Internet offerings means abandoning Regulation D's prohibition on general solicitation and general advertising. Commentators have long argued that the general solicitation and advertising ban is defective or unnecessary.[356] The SEC did, in fact, remove this proscription in Rule 504 offerings for a seven-year period during the 1990s only to reinstate it because of renewed concerns about fraud.[357] There are a number of potential benefits associated with open websites and the enhanced information they can provide.[358] Moreover, the general solicitation and advertising restriction is out-of-step with current communication norms, business practices, and lifestyles.[359] The SEC has thus far been reluctant to take an aggressive view on abandoning this dated proscription, despite its uncertain purpose and effect.[360] We have determined that abandoning

---

at 103 (noting that "[t]o be eligible for Regulation A, an issuer cannot be a reporting company under the 1934 Act. The point of this requirement is apparently to force public offerings by larger, 1934 Act companies onto either S Forms or SB Forms, with their more extensive disclosure requirements." (footnotes omitted)). While this limitation may not be essential, *see* Bradford, *supra* note 4, at 77–78, we offer it as part of the mix of attributes to be considered by rule-makers. We believe that regulatory costs are saved by varying the new exemption little from the existing Rule 504 exemption.

356. *See, e.g.*, Patrick Daugherty, *Rethinking the Ban on General Solicitation*, 38 EMORY L.J. 67, 70 (1989) (contending that the general solicitation and advertising ban is "unconscionably vague" and "broader than it should be"); Sjostrom, *supra* note 269, at 4 ("[T]here is no strong ideological foundation for the ban"); Yadley, *supra* note 282, at 82. ("Permitting greater use of general solicitation is not likely to diminish consumer protection or open the floodgates to fraud.").

357. *See* Bradford, *supra* note 145, at 19 (summarizing the relevant history); Campbell, *supra* note 146, at 97 n.92 (same). Professor Hazen, however, is very wary of delinking general solicitation and advertising from significant, substantive mandatory disclosure obligations. *See* Hazen, *supra* note 7, at 2, 10, 14, 19, 20. We understand the need for this linkage in the context of a private offering exemption under Section 4(2) and the related safe harbor under Rule 506 of Regulation D. However, we are concerned with requiring this linkage for all Section 3(b) offerings, regardless of offering size and other characteristics and terms.

358. *See* Bradford, *supra* note 4, at 78–79.

359. *See* Olufunmilayo B. Arewa, *Securities Regulation of Private Offerings in the Cyberspace Era: Legal Translation, Advertising and Business Context*, 37 U. TOL. L. REV. 331, 362 (2006) ("The SEC view of general solicitations and advertising is based on assumptions about the distribution of information that are inconsistent with post-Internet era standards of information dissemination and business practice.").

360. Professor Don Langevoort notes this puzzle in one of his articles:

> [W]e must ask why general solicitations are barred in the first place—something on which the Commission has never been particularly forthcoming. One possibility is simple concern for the statutory language, which speaks in terms of non-public *offerings* rather than sales. But there is no obvious reason why "offerings" must necessarily be given a meaning that precludes public advertising or mass mailings, and the statutory restraint concern is no longer applicable at all

general solicitation and advertising prohibitions is appropriate. However, to the extent that the ban on general solicitation and advertising actually limits investor risk or fraud, we believe it is important to place a renewed focus on addressing those values through other, more direct, investor and market protections.

One way to constrain investment risk is to limit an investor's exposure to losses in a particular enterprise.[361] This approach is a form of substantive regulation and is more than a bit patronizing. A per-investor cap curtails investor freedom by cutting the investor off from unacceptable losses before they occur, much in the same way that one might cut off a partygoer from inebriation by limiting the number of drinks that she may have. We have very mixed feelings about this aspect of our proposal.

As a result, we recommend limiting any per-investor cap to those investors who are not accredited or sophisticated. The logistics of implementing this type of hybrid requirement in crowdfunded offerings will not be trivial, and the costs may well exceed the benefits. In that event, a per-investor cap would be an unwise regulatory element. But those costs and benefits should be weighed in light of an overall proposal for a crowdfunding exemption. The implementation of a per-investor cap represents a heavy-handed form of investor protection, and if it is adopted, offering processes will need to be redesigned or modified. On the other hand, the nature of the crowdfunding market is such that it may attract participants who are intent on abusing the privilege of open solicitation and advertising.[362] Just as Professor Margaret Sachs suggests reaching out to protect the "least sophisticated investor" from fraud in inefficient markets,[363] it may be appropriate in these early stages of crowdfunding development to extend extraordinary protection to the unsophisticated and

---

in light of the Commission's new exemptive authority under the Act, which allows it to eliminate any statutory restrictions it wishes. Another possibility is that the prohibition is designed to protect the unsophisticated investor who might be tempted by the promotion into misrepresenting his or her qualifications in order to take part in the deal. That paternalistic concern is strained on its face; elsewhere, the Commission has recognized that since no prequalification procedures are fail safe, all we should require is *reasonable belief* by the issuer in the offeree's qualifications.

Donald C. Langevoort, *Angels on the Internet: The Elusive Promise of "Technological Disintermediation" for Unregistered Offerings of Securities*, 2 J. SMALL & EMERGING BUS. L. 1, 24 (1998) (footnotes omitted); *see also* Cohn & Yadley, *supra* note 144, at 36–42.

361. *See* Bradford, *supra* note 4, at 66–68 (discussing potential exemptions that limit investor losses to a "tolerable amount").

362. *See supra* Part IV.A.2; Hazen, *supra* note 7, at 20.

363. *See* Sachs, *supra* note 279.

unaccredited investors in a crowdfunded offering by limiting their capacity to invest in crowdfunding interests ex ante.[364]

In a similarly over-protective way, restricting the resale of crowdfunding interests may help constrain fraud, which may be more likely to occur in resale markets.[365] An investor who buys and sells securities in a resale market may find himself or herself attenuated from an accurate and complete source of information about the crowdfunded venture or the crowdfunding interest being offered. Typically, crowdfunding websites are designed to attract interest in primary offerings conducted over a short, defined period. Investors advance funds to a crowdfunded venture in order to achieve a specified funding target by a date certain.[366] Currently, most crowdfunding websites are not built to serve as markets for secondary offerings or even as hosts for ongoing disclosures that might support an appropriate secondary market. SEC endorsement of secondary trading in crowdfunding interests is unlikely in the absence of a reliable means for market participants to obtain current information about those interests and the crowdfunded venture.[367]

In practice and in theory, information is very important to investor protection and market integrity. Investors typically will not fund unknown risks.[368] Moreover, the regulation of securities offerings provided under the

---

364. *See* Hazen, *supra* note 7, at 20 (raising investor protection concerns based on the possible nature of crowdfunding investors). An interesting question is whether the per-investor cap should be an aggregate cap for all of a single investor's crowdfunding investments (perhaps together with investments made by affiliates and associates) or whether the cap should only apply to investments in a particular crowdfunded venture. Professor Steve Bradford posits that the former is more reasonable. *See* Bradford, *supra* note 4, at 76. We suggest otherwise, *see supra* note 340 and accompanying text, but understand and appreciate Professor Bradford's concern and would reassess this aspect of a possible exemption in light of other attributes of a specific exemption proposal.

365. In reinstating the prohibition on general solicitation and advertising under Rule 504 in 1999, the SEC noted concern about fraud in the trading markets for securities offered and sold by issuers under Rule 504. *See* Bradford, *supra* note 145, at 19. The risk may have been overstated or given undue effect in the SEC's decision-making, however. *Id.;* Campbell, *supra* note 146, at 97 n.92.

366. *See supra* Part II.B.3 (describing this funding model in the context of the commonality element of the *Howey* test).

367. This apparently was a further concern of the SEC in its decision to reinstate the general solicitation and general advertising requirement to Rule 504. *See* Bradford, *supra* note 145, at 19 ("The SEC believed that . . . the lack of widely-distributed public information about companies making Rule 504 offerings, and the freely tradable nature of Rule 504 securities may have exacerbated the opportunities for microcap fraud."). We note that it is difficult for issuers in this context to constrain resale transactions. However, any regulatory solution should address the manner in which investor violations of any resale prohibition impact the issuer's exemption. *See* Cohn & Yadley, *supra* note 144, at 54–58 (raising this issue).

368. This principle can be illustrated simply in a basic discussion of business finance.

Securities Act reflects the semi-strong version of the efficient capital markets hypothesis.[369] As a result, the Securities Act requires the disclosure of investor-significant information through mandatory disclosure rules (in the form of prospectuses, offering circulars, and otherwise in registration exemptions) and through fraud-protection rules that call for the disclosure of material information where there is a duty to disclose.[370] Neither Congress nor the SEC has deemed mandatory disclosure or fraud protection to be a sufficient regulatory tool without more.[371] Consistent with these

---

Professor Heminway has been known to begin her first Securities Regulation class of the semester with a simple statement and question that illustrate this basic point. She says something like: "I have started a business. Do you want to buy an interest in it?" Students hesitate, and when she asks why (to the extent they do not offer a reason), they respond with something akin to: "Well, before I put down my hard-earned money, I would like to know something about the business." Further discussion illuminates that they want to know about the finances and operations of the business, as well as the nature of the interests being offered and where the offering proceeds will go. A well-guided discussion can touch on all of the areas of mandatory disclosure in a Securities Act registration statement. Of course, these students have been assigned (and may have read) the introductory chapter of our casebook as background to this discussion.

369. *See* Robert A. Brown, *Financial Reform and the Subsidization of Sophisticated Investors' Ignorance in Securitization Markets*, 7 N.Y.U. J. L. & BUS. 105, 165 (2010) ("[M]any legal commentators, courts and lawmakers have cited the EMH as a guiding principle in articulating rules of the architecture of American securities markets. The Securities Act of 1933 and Securities Exchange Act of 1934, which preceded proceeded formal development of the theory, have at their heart a desire to disclose to investors unknown risks." (footnote omitted)); Michael W. Prozan & Michael T. Fatale, *Revisiting "Truth in Securities": The Use of the Efficient Capital Market Hypothesis*, 20 HOFSTRA L. REV. 687, 697 (1992) ("The Commission explicitly adopted the ECMH in its refinement of the registration process.").

370. *See supra* Parts III.A.1, 2.

371. In early presentations of this article to faculty audiences, we advocated mere reliance on antifraud rule protections, suggesting that (with a minimization of investor risk in other ways) mandatory disclosures in the form of SEC filings or investor information materials were not needed. After some push-back from those audiences, we rethought the issue and determined that minimal mandatory disclosures would best serve the policies underlying the Securities Act and support important related regulatory values. *See* Cohn, *supra* note 150, at 365 ("The impetus for disclosure in the nonregistered setting is compliance with exemption conditions and effective sanctions."). The exemption could (and perhaps should) expressly designate the required mandatory disclosures as, individually and collectively, a prospectus for the purposes of Section 12(a)(2), affording the purchasers of crowdfunding interests a Securities Act cause of action for material misrepresentations and omissions. Current law may already afford investors that right. *See infra* note 374. Professor Hazen, among others, favors giving crowdfunding investors this right of action. *See, e.g.*, Hazen, *supra* note 7, at 14, 22. Any plan to abandon mandatory disclosure must meet a high burden of proof, given its centrality to the federal securities regulation scheme.

> [A]t the same time that the Internet is increasing the impact of small business offerings, regulatory reform efforts . . . may be effectively transforming [Internet

regulatory objectives, and in furtherance of our foundational principles of enhancing informational transparency and fostering disclosure and enforcement standardization, we suggest that putative rule makers consider requiring issuers/offerors to (a) file with the SEC both an issuer registration form and an abbreviated brief notice of each offering and (b) include cautionary language and certain other limited disclosures on the crowdfunding website though which the offering is made.

The important issue in fashioning these filing and disclosure requirements is balancing the desired level of information against the costs of producing and disseminating that information.[372] Indeed, "[t]oo much complexity at the entrepreneurial level will . . . destroy the exemption's utility."[373] Accordingly, our suggested filing and other disclosure requirements are intended to be minimal but substantive. At the low dollar-value level of investment that issuers would be requesting and investors would be making under the exemption, most of the disclosure requirements for crowdfunding websites and crowdfunded ventures would be satisfied by complying with the anti-fraud protections afforded by Section 17 of the Securities Act and Section 10(b) of and Rule 10b-5 under the Exchange Act.[374]

In that spirit, we envision SEC filing requirements that are simple notice submissions—one or more forms that would represent a tailored version of either the Form D required for offerings under Regulation D[375] or

---

> offering] regulation . . . from a prophylactic disclosure structure to one that merely reacts to and combats fraud. The original promulgation of the federal securities laws was based on congressional perception that such a structure was an ineffective means of regulating the national securities markets. Proponents of regulatory reform need to explain why technological developments since the 1930s have rendered that perception obsolete.

Fisch, *supra* note 274, at 89.

372. Constraining costs to issuers is a foundational principle for our proposed regulatory solution, as set forth *supra* Part IV.B.1.

373. Bradford, *supra* note 4, at 68.

374. *See supra* notes 140–42 and accompanying text. We also note the probable application of Section 12(a)(2) of the Securities Act to prospectuses used in offerings exempt under Section 3(b), even in the event there is no express provision in the exemption itself. *See* Elliott J. Weiss, *Some Further Thoughts on* Gustafson v. Alloyd Co., 65 U. CIN. L. REV. 137, 152 (1996) ("[M]ost important . . . are . . . small-scale offerings made pursuant to section 3(b). . . . [A]t least where securities are sold to the public, section 12(2) applied pre *Gustafson*, and section 12(2) would continue to apply."); Natasha S. Guinan, Note, *Nearly a Decade Later: Revisiting Gustafson and the Status of Section 12(a)(2) Liability in the Courts—Creative Judicial Developments and a Proposal for Reform*, 72 FORDHAM L. REV. 1053, 1069 (2004) ("Section 12(a)(2) liability expressly attaches to Section 3 offerings by referring to: 'Any person who . . . offers or sells a security (whether or not exempted by the provisions of [Section 3]).'").

375. *See* 17 C.F.R. § 230.503 (2011). A PDF version of Form D is available at

the Small Company Offering Registration form.[376] For example, the issuer registration would include basic information about the crowdfunding website or crowdfunded venture, including information necessary for locating and notifying the filer and its relevant personnel in connection with monitoring and enforcement.[377] Similarly, the level of information required for the notice of each offering would be comparable to that required to complete a Form 144 under the Securities Act.[378] Logically, the amount of required disclosure in the company registration and offering notice would bear an inverse relationship to the aggregate size of the offering and the dollar value of the per-investor cap.[379] However, a certain minimal amount of information necessary for monitoring and enforcement would be required as a threshold matter.[380]

In addition, rule makers should consider requiring the inclusion of certain cautionary language and other disclosures on websites through which crowdfunded offerings of securities are made. Professor Steve Choi aptly summarizes the behavioral psychology basis for these types of requirements in securities regulation as a means of protecting investors:

> One possible method of correcting for behavioral biases is to provide corrective or cautionary information to investors. If investors view sales materials too optimistically, then providing the investors more sober

---

http://www.sec.gov/about/forms/formd.pdf.

376. The Small Company Offering Registration Form (Form U-7) is available at http://www.nasaa.org/industry-resources/corporation-finance/scor-overview/scor-forms/.

377. Although registration requirements will not enable the SEC, the Department of Justice, or the Federal Bureau of Investigation to find every Internet fraudster, we offer it as a way to ameliorate the effects of the faceless, opaque, remote nature of the Internet. *See* Fisch, *supra* note 274, at 81 ("The power of the Internet to transcend jurisdictional boundaries suggests . . . that it may be more difficult for victims and regulators to trace the source of fraudulent offers and obtain legal recourse against wrongdoers." (footnote omitted)).

378. *See* 17 C.F.R. § 230.144(h). A PDF version of Form 144 is available at http://www.sec.gov/about/forms/form144.pdf.

379. *Cf.* C. Steven Bradford, *Expanding The Non-Transactional Revolution: A New Approach To Securities Registration Exemptions*, 49 EMORY L.J. 437, 449 (2000) ("Intermediate disclosure rules that do not provide the full benefit of registration, but also have lower compliance costs, could be economically efficient for all but the smallest offerings. An incremental system in which the level of investor protection increases as the size of the offering increases could make sense." (footnote omitted)).

380. We acknowledge that standardized disclosures have a cost to the crowdfunded venture and crowdfunding website. We believe that these costs are not significant, but we may be wrong in this regard. *See* Bradford, *supra* note 4, at 84–85. A more detailed cost-benefit analysis can be made as proposals develop through engagement among the relevant constituencies. If disclosures necessary to appropriate risk reduction are not cost-effective for crowdfunding websites, crowdfunded ventures, and crowdfunding investors, then the exemption we envision here should be abandoned. The adoption of an exemption that will seldom, if ever, be used is a waste of regulatory time and effort.

> materials on the issuer's business, properties, and financial health may, in theory, help overcome their overoptimism. Additional information may serve to educate investors about the potential pitfalls they face in investing . . . or, alternatively, caution the investors to take extra care in their investment decisions. If investors are capable of learning, then mandatory disclosure and legends may work to educate investors, reducing their behavioral biases.[381]

In a crowdfunding context, these required disclosures could be made on the web page for each crowdfunded venture. The most critical information could also be displayed for would-be investors to acknowledge as a condition to the crowdfunding website accepting their investment funds (through, e.g., something akin to a click-wrap agreement,[382] implemented through a pop-up window with a check-the-box requirement). Warning and advisory legends are already used in disclosure rules and exemptions under the Securities Act.[383] Although many may read over a standard cautionary legend without heeding or reacting rationally to its content,[384] we believe that those investing in crowdfunding interests should at least have notice that their entire funded amount is at risk, that the probability of any return is remote, and that their interests are illiquid. Discussions among regulators and those in the industry will help decide whether and how additional cautions should be issued. This approach represents a minimal and inexpensive means of promoting investor protection and fosters transparency and standardization.[385]

---

381. Stephen J. Choi, *Behavioral Economics and the Regulation of Public Offerings*, 10 LEWIS & CLARK L. REV. 85, 116 (2006) (footnote omitted).

382. *See generally* Francis M. Buono & Jonathan A. Friedman, *Maximizing the Enforceability of Click-Wrap Agreements*, 4 J. TECH. L. & POL'Y 13 (1999), *available at* http://jtlp.org/vol4/issue3/friedman.html ("A 'click-wrap agreement' is an agreement, formed entirely in an online environment such as the Internet, which sets forth the rights and obligations between parties.").

383. *See* 17 C.F.R. § 230.134(b)–(d) (requiring certain statements in communications used after filing a registration statement that may otherwise be prospectuses within the meaning of the Securities Act); 17 C.F.R. § 230.135(a)(1) (requiring in a communication used before the filing of a registration statement "a statement to the effect that it does not constitute an offer of any securities for sale" so that the issuer will "not be deemed to offer its securities for sale through that notice"); 17 C.F.R. at § 230.163(b)(2) (requiring issuers to state that they may file a registration statement with the SEC and directing potential investors to read the prospectus included in that filing); 17 C.F.R. § 230.433(c)(2) (requiring issuers to state that they have filed a registration statement and directing potential investors to read the prospectus included in that filing).

384. For a pithy critique of legending requirements (focused on those in connection with the SEC's 2005 offering reforms), *see* Choi, *supra* note 381, at 118–19.

385. Before implementing the legending requirement, however, we suggest that the SEC study its efficacy. *See id.* at 128 (suggesting four ways in which the SEC should clarify and make explicit the assumptions about investors on which its regulatory proposals rest). Consistent with our foundational principles of constraining regulatory costs and minimizing

Transparency and standardization, as well as (potentially) investor protection and fraud prevention, also may be promoted by mandating certain simple disclosures on the crowdfunding website relating to the crowdfunding website, the crowdfunded ventures, the interests being offered, the way in which the offering is being conducted, the ongoing role of the crowdfunding website after investments are made, and any follow-on ministerial services that will be rendered, such as delivery of investor funds to the crowdfunded venture, monitoring of the crowdfunded venture's operations and financial data, and collection and distribution of profit-sharing or revenue-sharing amounts to investors. The major disadvantages of this type of disclosure requirement are its cost and potential to stifle efficient, desirable innovation.[386] Again, however, we posit that the required disclosures could be minimal given the relatively low amount at risk, if low caps on both the aggregate offering price and per-investor funding are instituted. More disclosure is not necessarily more protective to investors.[387] We suggest an assessment of investor needs, a review of industry best practices, and a touch of aspirational imagination to identify the nature and extent of appropriate, efficient, and efficacious mandatory disclosure requirements in the crowdfunding context. We would hope that the result would be a competition for investors based on, among other factors, the quality of the crowdfunding website's disclosures.

Neither the suggested process for regulatory change nor the substantive recommendations we outline in this article addresses market regulation issues under the Exchange Act, investment advisory issues under the Investment Advisors Act of 1940, the interaction of crowdfunding with state securities (or "Blue Sky") rules, or the inherent cross-border nature of

---

costs to issuers and investors, disclosure burdens should not be added unless their benefits clearly exceed their costs. *See id.* at 119 (noting that the SEC likely bases its decisions to impose legends on an "*ad hoc* basis"); *id.* at 122 ("[T]he SEC implicitly makes assumptions about how investors behave"). We acknowledge the possibility that standardization, if taken too far, can dampen positive entrepreneurial innovation. Bradford, *supra* note 4, at 85.

386. Professor Alan Palmiter expresses the disadvantages well and advocates that issuers be permitted to choose the level of disclosure they offer:

> The Securities Act often compels issuers to disclose and warrant more than investors are willing to pay for, driving issuers to avoid mandatory disclosure or to choose other financing techniques. Disclosure choice in securities offerings promises to expand the methods and reduce the costs of capital formation by aligning disclosure (both its contents and methods) with actual investor information demands, not legislative or administrative assumptions.

Alan R. Palmiter, *Toward Disclosure Choice In Securities Offerings*, 1999 COLUM. BUS. L. REV. 1, 86 (1999).

387. *See* Choi, *supra* note 380 ("While more disclosure into the market may help some investors, the increased information may simply cause others to fall further into the traps of overconfidence and overoptimism.").

Internet securities offerings (including crowdfunded offerings).[388] We leave federal market regulation and investment advisory issues to another commentator or another article.[389] As to state securities law interactions, we believe that the SEC is best positioned to administer and enforce crowdfunding regulation (given the interstate and international nature of crowdfunding), which would require federal preemption of state regulation.[390] Finally, as to the globalization of securities markets and transactions, we join the growing chorus of voices that urge continued consideration of a more coherent approach to international regulation and enforcement in an increasingly global transactional world.[391]

---

388. *See supra* notes 7, 218, 237, 250, 341 (noting these exclusions as well as the exclusion of other regulatory schemes, such as gambling and charitable solicitation regulation, from treatment in this article). We also do not address the precise status of crowdfunding websites for purposes of the overall regulatory scheme under the Securities Act. *See supra* note 341. Any crowdfunding regulatory solution, regardless of whether it follows any or all of the elements of the proposal we set forth in this article, should address the issue of the regulatory status of crowdfunding websites under all applicable securities laws. Other authors already are approaching some of these issues. *See* Bradford, *supra* note 4, Hazen, *supra* note 7; Pope, *supra* note 215.

389. *See* Bradford, *supra* note 4, at 32–51. The need for improved regulation of market professionals extends beyond the crowdfunding context. *See, e.g.*, Jennifer J. Johnson, *Private Placements: A Regulatory Black Hole*, 35 DEL. J. CORP. L. 151, 191 (2010) ("[A]s the recent financial crisis has demonstrated, there is much room to improve the regulation of institutions that intermediate between individual investors and the securities markets. Many scholars believe that the SEC should increase its oversight of these intermediaries, such as investment advisors and broker-dealers." (footnote omitted)).

390. *Accord* Bradford, *Peer-to-Peer Lending*, *supra* note 7 ("A federal exemption that does not preempt state law isn't going to accomplish much."). This is not a new suggestion. Scholars have argued for preemption in connection with existing registration exemptions. *See, e.g.*, Bradford, *supra* note 145, at 33–34 ("Small businesses should receive the same consideration that the 1996 Act gave public companies: the states should be preempted from requiring the registration of Rule 504 offerings."); Campbell, *supra* note 146, at 106–110, 119 (describing the high cost of state securities compliance for small businesses using Regulation A offerings and concluding that "[t]he best way to eliminate state interference is for Congress to expand the preemption of NSMIA to include securities issued under Section 3(b) of the 1933 Act."); Rutheford B Campbell, Jr., *The Insidious Remnants of State Rules Respecting Capital Formation*, 78 WASH. U. L. Q. 407, 413–33 (2000) (arguing generally for federal preemption of state securities law regulation of capital formation). *But see* Johnson, *supra* note 389, at 192 (arguing for a return to state regulation "of Rule 506 private placements by private entities to largely retail investors."). Unlike our proposal for handling the Securities Act registration exemption, this change would require congressional action. *See* Hazen, *supra* note 7, at 16–17.

391. *See, e.g.*, Chris Brummer, *Post-American Securities Regulation*, 98 CALIF. L. REV. 327 (2010) (noting the need for, barriers to, and prospects for international securities regulation); Eric C. Chaffee, *Contemplating the Endgame: An Evolutionary Model for the Harmonization and Centralization of International Securities Regulation*, 79 U. CIN. L. REV. 587 (2010) (arguing for reform, harmonization, and centralization of securities regulation); Stephan J. Choi & Andrew T. Guzman, *Portable Reciprocity: Rethinking the International*

V. CONCLUSION

Crowdfunding is an exciting, dynamic, inclusive capital formation model for small businesses (and, in some present and envisioned future forms, for larger business ventures). However, crowdfunded ventures and crowdfunding websites that offer profit-sharing interests to funders violate Section 5 of the Securities Act when they offer or sell those interests without registration or compliance with an applicable exemption.[392] Investor protection and the perception of fair and honest investment markets—key policies underlying the Securities Act—are sources of concern as the crowdfunding market rapidly develops in the absence of a clear regulatory framework or response. With the thought of harnessing crowdfunding's positive attributes and minimizing its potential negative characteristics, this article suggests a conservative approach to regulating crowdfunding through a new SEC registration exemption under Section 3(b) of the Securities Act.

Yet, as this article amply illustrates, the exemption process will not be simple. It will require a delicate balancing of interests among the SEC, industry participants, and investors. It will involve parallel action by Congress. This is, of course, not new. The debate over crowdfunding regulation illustrates the classic tension between regulatory and market-based solutions to perceived and actual market failures. SEC Commissioner Troy Paredes states the basic issues well:

> Government intervention in securities markets to put information in investors' hands and to protect investors against corporate abuses serves a distributional goal by protecting investors against losses. Such government intervention also serves the larger goal of promoting capital formation and more efficient and liquid securities markets in that investor protection regulation can shore up investor confidence in the integrity of securities markets. Sometimes, though, increased investor protection, such as through more mandatory disclosure and more aggressive SEC oversight and enforcement, can impede market participation and thus undercut the capital formation process and the efficiency and liquidity of securities markets.
>
> This tension drives the cost-benefit analysis of regulating securities markets. Regulatory systems that allow for flexible, dynamic financial markets inevitably come at the risk of investor loss, fraud, and corruption. Regulators have to exercise restraint and allow for misconduct and abuse of investors because, at some point, investor protection overburdens

---

Reach of Securities Regulation, 71 S. CAL. L. REV. 903 (1998) (proposing "portable reciprocity" and recommending "a regulatory regime that focuses on regulatory competition and gives issuers and investors the ability to choose the law that governs their transactions.").

392. *See supra* notes 133–35 and accompanying text.

> financial markets. The question of when it becomes too costly for the government to protect investors is a fundamental challenge of securities regulation.[393]

The determinations that must be made to resolve this tension are all-the-more difficult in the crowdfunding context because of the way in which crowdfunding interfaces directly with rapidly changing technology, state securities regulation, and globalism. We do not have all of the answers to the questions that may be raised about regulating crowdfunding. However, we do believe that it is important that the process and outcome of crowdfunding regulation strike an appropriate balance that both fosters crowdfunding's promise—as a means of raising investment funds for small businesses and allowing individual retail investors to access a user-friendly business finance market—and supports policies and values central to both securities regulation and crowdfunding's potential ongoing role in small-business capital formation. In that spirit, this article is designed to contribute positively to the regulatory debate.

393. Troy A. Paredes, *On the Decision to Regulate Hedge Funds: The SEC's Regulatory Philosophy, Style, and Mission*, 2006 U. ILL. L. REV. 975, 1006.

TABLE 1

U. S. CROWDFUNDING COMPANIES WITH PROFIT-SHARING OR REVENUE-SHARING ARRANGEMENTS[a]

Promoter; Nature of Project Funded	Funder Pseudonym; Attributes	Unit of Interest; Price per Unit; Max # of Units; Max Aggregate Value	Financial Return to Funder	Other Funder Benefits	Contributors to Profit Generation	Financial Return to Promoter[b]	Financial Return to Project
Appbackr www.appbackr.com Development of mobile apps to be sold by the promoter on Apple's App Store	Backr	No units; Backr purchases copies of apps in bulk from promoter, which will then be held in a queue as they are sold on the App store at retail	As apps are purchased at retail, the profit is split between the funder and the developer of the app for the amount of apps that the funder purchased	None	Funder pays to have apps listed on App Store; promoter lists the apps on the App Store; developer responsible for the app's retail success	Promoter's profit share is roughly 13% of retail revenue	Developer retains all ownership rights in the app (only sells copies to the Backr); developer typically receives 45% of sales revenue for concept apps and 35% for apps already sold on the App Store
CinemaShares[c] www.cinemashares.com Movies; allows for the purchase of fully listed (NASDAQ), publicly traded stock in a movie (currently, the Web site is not selling stock)	Investor; must register as a member of the site	Stock; $20/share; minimum funding target of $10 million	Profit-sharing based on the number of shares owned	DVD, pass to observe filmmaking, access to directors and producers of the film	Producers and directors of the movie; promoter underwrites the direct public offering on NASDAQ	Not disclosed	Not disclosed

[a] The list of websites in this Appendix was compiled on August 23, 2011. This table was updated on December 28, 2011 to verify the accuracy of the listed websites.
[b] "Promoter" means the crowdfunding website.
[c] CinemaShares.com is a licensee of MediaShares.com, which utilizes the same business model.

U. S. CROWDFUNDING COMPANIES WITH PROFIT-SHARING OR REVENUE-SHARING ARRANGEMENTS

Promoter; Nature of Project Funded	Funder Pseudonym; Attributes	Unit of Interest; Price per Unit; Max # of Units; Max Aggregate Value	Financial Return to Funder	Other Funder Benefits	Contributors to Profit Generation	Financial Return to Promoter[d]	Financial Return to Project
MicroVentures www.microventures.com Small businesses (predominantly those that focus on technology, finance, and healthcare)	Investor	Stock; investments are typically between $250 and $5000; sold as a private offering	Equity ownership	None	Principal	Principal must pay a $100 application fee, $250 screening fee if selected, and 10% of offering price to promoter	Principal retains residual ownership
ProFounder www.profounder.com Entrepreneurial business ventures (must be a U.S. for-profit business)	Investor; limited to individuals invited by the principal; maximum number of investors limited by state laws of the investors' residence (normally 35)	Pledges; minimum of $100 per pledge; utilizes Rules 504 and 506 of Regulation D	Share of quarterly revenues of the business (based on terms established by principal)	None	Principal; promoter creates the website to market the business and may perform administrative duties at the election of the principal	Promoter receives $100 at the beginning of fundraising and, if the business is successful, performs administrative work for a fee of $1,000	Principal retains all ownership in the business, but will only receive the revenue not distributed to funders
Quirky www.quirky.com Innovative product ideas which are manufactured by promoter and then sold via an online store or at	Community member	Influence; these entitle the supporter to share of revenue but are not sold at a specific price per share; no maximum # of people who influence; no	30% of revenue (and 10% of indirect retail revenue) generated by sale of the product is split among the Quirky community and distributed pro rata based on the	None	Principal; promoter (who actually engineers, manufactures, sells, and markets the product); often other members influence the project by	Principal pays a $10 submission fee; promoter retains 70% of revenue from online sales (90% of revenue from retail sales are	About 35% of the 30% of revenues shared in the Quirky community goes to the principal

[d] "Promoter" means the crowdfunding website.

retail		maximum value	amount of "influence" each individual contributed to a project; "influence" includes submitting the idea, financially supporting the idea, voting on the idea, refining the idea, etc.		contributing suggestions or comments to the project	allocated between promoter and retailers)	
SeedUps www.seedups.com Provides start-up capital for entrepreneurs	Investor; must be "qualified high net worth / sophisticated investor"	Maximum aggregate amount of $250,000	Equity share in the business	None	Principal responsible for success of the business; Promoter matches the funder with a specific project using its "matching engine"	Unspecified matching fees from both principal and investors	Residual equity ownership
33needs[c] www.33needs.com Entrepreneurships with a social/humanitarian mission	Investor	Generally, there are no units for the investment; however, funders can earn Impact Points ($1 = 1 Impact Point)	Investment dollars are traded for "financial and emotional goodies," dependent upon the project	Impact points and investments entitle the funder to various nonfinancial "rewards," in addition to the altruistic emotional benefits	Promoter screens the ventures and provides a medium for funding, but success ultimately depends on the principal	5% of funding target	Principal retains all ownership and intellectual property rights

[c] Because 33Needs now requires membership to access the website, this information is current as of August 2011.

U. S. CROWDFUNDING COMPANIES WITHOUT PROFIT-SHARING OR REVENUE-SHARING ARRANGEMENTS

Promoter; Nature of Project Funded	**Funder Pseudonym; Attributes**	**Unit of Interest; Price per Unit; Max # of Units; Max Aggregate Value**	**Financial Return to Funder**	**Other Funder Benefits**	**Contributors to Profit Generation**	**Financial Return to Promoter**	**Financial Return to Project**
Chip In www.chipin.com Allows individuals to raise money for any personal cause or fundraising effort	Contributor	None	None	None	Promoter provides the interface to promote the project	None	Funds for the non-profit project are raised, but there are no returns to the project
Citizen Effect www.citizeneffect.org Philanthropic projects	Citizen Philanthropist	None	None	None	Principal is responsible for carrying out the project	None	Project is funded, but there are no returns to the project
GlobalGiving www.globalgiving.org Provides donations to philanthropic projects	None	No units; minimum individual donation of $10; minimum aggregate amount of $250	None	None	Promoter and "project sponsors" choose the projects for funding; "project leaders" and implementing organization responsible for the project's success	15% of funds raised	Funds for the non-profit project are raised, but there are no returns to the project
IndieGoGo www.indiegogo.com Various creative or entrepreneurial projects and causes	Fans	No units; individuals simply pledge money and there are no explicit limits on aggregate amount	None	"VIP Perks" chosen by the project creators (all appear to be non-financial)	Principal: promoter provides various social media and marketing tools	4% of funds raised, but if funding target is not reached, promoter may retain 9% of funds raised (if the "flexible funding" method is chosen)	Principal retains 100% ownership

Kachingle www.kachingle.com Donates a portion of $5 per month to any group of participating websites (selected by the funder) in proportion to the number of days that the funder visits each website during a month	Kachingler	No unit; individuals select favorite sites to "kachingle" each month; limited to $5 per month	None	None	None; Principal (the participating website) receives a payment each month as long as the funder continues to "kachingle" the site	15% of all funds raised are retained by promoter for use of PayPal and as commission	Participating site is paid 85% of the portion that is to be allocated to that based on the funder's visits each month
Kickstarter www.kickstarter.com Artists' various projects are funded	None	No unit; individuals pledge money and there are no limits on the aggregate amount raised	None; almost anything can be given as an reward, but "investment and loan solicitations are forbidden"	Various rewards determined by the artist; these can range from $1 to $10,000 in value	Artist	5% of the total amount funded once target is met	Artist retains all ownership of the project and is entitled to all profits
Kiva www.kiva.org Small businesses in undeveloped countries	None	No unit; no explicit limits on aggregate amount and individual funders can fund from as little as $25 and as much as the entire requested amount	None; lenders get their loan back with no interest	None	Entrepreneur; "field partner" (microfinance institutions): field partner keeps the interest on the loan to cover costs	None	Entrepreneur keeps profits after loans are repaid
Peerbackers www.peerbackers.com Entrepreneurial projects and businesses	Backer	None; no limits on amount funded	None	Various rewards at the principal's discretion, but cannot include financial returns	Principal responsible for success of the business	5% of amount funded	Principal retains all ownership and profits

U. S. CROWDFUNDING COMPANIES WITHOUT PROFIT-SHARING OR REVENUE-SHARING ARRANGEMENTS

Promoter; Nature of Project Funded	Funder Pseudonym; Attributes	Unit of Interest; Price per Unit; Max # of Units; Max Aggregate Value	Financial Return to Funder	Other Funder Benefits	Contributors to Profit Generation	Financial Return to Promoter	Financial Return to Project
RocketHub www.rockethub.com Various artists' projects	Fueler	RocketFuel; $1/Liter of RocketFuel; no explicit limits on aggregate amount	None	Various items at the artists' discretion	Artist	8% of funded amount	Artist keeps 100% ownership and profits
40Billion www.40billion.com *Entrepreneurial businesses*	None: funders must be solicited privately (i.e., invited by principal or another funder that was invited by principal)	Three types of funding: (1) gift/contributions are limited to $25 - $10,000 per person; (2) direct loans are limited to $25 - $10,000 per person; (3) commercial paper is limited to $100 - $1,000,000 per person: maximum funding target $99,000, unless commercial paper is used, in *which case the limit is* $1,000,000	Interest on direct loans or commercial paper	None beyond helping a friend or family member, since all funders must be invited	Promoter structures the debt financing, markets, and provides resources for connection to other businesses; principal responsible for success of business	There are currently no start-up fees; fees of $9.99 - *$49.99* per month to extend funding period	Principal retains 100% ownership

FOREIGN CROWDFUNDING COMPANIES WITH PROFIT-SHARING OR REVENUE-SHARING ARRANGEMENTS

Promoter: Nature of Project Funded	**Funder Pseudonym; Attributes**	**Unit of Interest; Price per Unit; Max # of Units; Max Aggregate Value**	**Financial Return to Funder**	**Other Funder Benefits**	**Contributors to Profit Generation**	**Financial Return to Promoter**	**Financial Return to Project**
Buzzbnk www.buzzbnk.org Debt financing for social projects	Backer	BITs (loan units); no explicit limits	Backers' loans may entitle them to either: (1) share of revenue for the term of the agreement; (2) interest for the term of the agreement; or (3) repayment of principal only	Additional benefits as determined by the principal	Principal	£25 registration fee and an administrative fee of 5% of funds raised	Principal retains 100% ownership
Crowdcube www.crowdcube.com Entrepreneurial businesses	None; limited to UK residents age 18 or over	None; minimum of £10 per person; maximum of £5,000 per person unless that person qualifies as a High Net Worth Individual or Sophisticated Investor; minimum aggregate amount of £5,000; no maximum aggregate amount	Funders receive a share of profits by virtue of equity ownership, as determined by the principal	Rewards (usually gifts or services of the business) at the principal's discretion	Success of business depends on principal	£250 listing fee (waived for a limited time), 5% of funds successfully raised, and legal fees of £1750 for each completed investment	Principal retains residual interest in equity not distributed to investors (determined before the project can seek funding)

U.S. CROWDFUNDING COMPANIES WITH PROFIT-SHARING OR REVENUE-SHARING ARRANGEMENTS

Promoter; Nature of Project Funded	Funder Pseudonym; Attributes	Unit of Interest; Price per Unit; Max # of Units; Max Aggregate Value	Financial Return to Funder	Other Funder Benefits	Contributors to Profit Generation	Financial Return to Promoter	Financial Return to Project
Investiere www.investiere.ch/ Provides capital for Swiss start-up companies; sites says it is not crowdfunding, but more of a venture capital business model	Investor	Minimum investment is usually around $250 per investor; maximum is usually $10,000 per investor; two investment models are (1) equity, and (2) mandatory convertible debt	Funder receives an equity share in the company	None	Promoter performs due diligence and selects the funded ventures; promoter also helps manage financial aspects of the business after the financing round; principal is otherwise responsible for growth and success of the business	6.5% of funds raised	Residual equity interest
SellaBand www.sellaband.com Various music projects, mainly albums	*Believer*	Part: price varies, but typically €10/Part; no maximum # of Parts; maximum amount of €250,000	Revenue sharing (at artist's discretion)	Music downloads, CDs, t-shirts	Artist is responsible for success of album; promoter contributes to the artist's fundraising by offering clips of songs on the website	Fee of 15% of amount funded for "providing the platform and expertise in the music industry"	Artist retains all ownership and is entitled to all profits (less any revenue-sharing agreement with funders)
SonicAngel www.sonicangel.com Music projects	Angel	FanShares; €10/FanShare; maximum of 100 FanShares (€1,000) per individual	Revenue sharing, usually around 25-30% of net sales	Name published on website, album download, and various "Behind the Scenes" perks	Artist: promoter places the music on iTunes (and other online music platforms) and in retail stores, and has connections with various record companies	Promoter takes zero commission on funders' contributions; however, promoter enters into agreement with artist giving promoter recording, managing, and publishing rights	Artist and promoter share the remaining 70-75% of net sales from the funded project per the terms of the agreement; artist must relinquish some control to the promoter under the "Full Rights Contract"

Spanner Films (*Age of Stupid*) spannerfilms.net Raised over £850,000 for a film on global warming	None	Shares; varied between £500 and £35,000	Profit share for those who donated more than £5,000	Those who contributed less than £5,000 got their names in the credits	Director/producers (who were also the promoters) were ultimately responsible	Promoters were entitled to profits remaining after pro-rata distributions are made to funders	Remaining profits
VenCorps www.vencorps.com Promoter provides $25,000 in venture capital to startups picked by the site's members	Facilitators, who provide peer reviews and advice and ratings on different startups listed on the site; Funders, who can invest along with promoter	Points, which are earned by participating in reviewing and backing (by voting) a particular startup; no explicit limits	Funders receive an equity interest; facilitators receive no equity interest (only points)	Points earned by facilitators can be redeemed for various non-financial goods and services	Funders and facilitators choose which startup should be funded; principal and promoters manage the company	Promoter receives equity in exchange for the investment	Principal retains residual ownership of the company

FOREIGN CROWDFUNDING COMPANIES WITHOUT PROFIT-SHARING OR REVENUE-SHARING ARRANGEMENTS

Promoter; Nature of Project Funded	Funder Pseudonym; Attributes	Unit of Interest; Price per Unit; Max # of Units; Max Aggregate Value	Financial Return to Funder	Other Funder Benefits	Contributors to Profit Generation	Financial Return to Promoter	Financial Return to Project
FansNextdoor[6] www.fansnextdoor.com Various creative projects	Fan	None: no explicit limits on maximum # or value of shares	None	Rewards are chosen by the principal and usually reflect the amount contributed	Principal; fans are encouraged to promote the product online and offline	Not disclosed	Principal retains 100% ownership; promoter retains a license
Pledge Music www.pledgemusic.com Music albums	Fan	None	None	Incentives include a copy of the album, merchandise, special events, etc., at the discretion of the artist	Artist; promoter contacts the artist's mailing list to let them know about the project	15% of funded amount	Artist retains 100% ownerships
Pozible http://pozible.com.au Various creative projects: developed for artists, musicians, filmmakers, journalists, designers, entrepreneurs, inventors, etc.	None	None; no explicit limits on maximum # or value of shares	None	Principal has the ability to choose different tiers of rewards (none appear to be monetary)	Principal is solely responsible for the project's success	5% of funded amount for those who were invited to post the project by promoter; 7.5% for everyone else	Principal retains 100% ownership
Ulule www.ulule.com Allows individuals to raise money for any personal cause or fundraising effort (most projects are those that enhance social welfare)	None	None	None	Rewards (nonfinancial) determined by the principal, typically related to the project	Principal	5-8% commission on the collected amount of successful projects	Principal retains 100% ownership of project

[6] The site was recently restructured and no longer has information under the "How it Works" page. This information provided is current as of February 2011.

SALES TO GRANTOR TRUSTS: A CASE STUDY OF WHAT THE IRS AND CONGRESS CAN DO TO CURB AGGRESSIVE TRANSFER TAX TECHNIQUES

JAY A. SOLED* & MITCHELL GANS**

Sales to grantor trusts produce magnificent transfer tax savings. Such savings raise an important policy question: What can the IRS and Congress each do to eliminate this and other transfer tax savings devices that erode the transfer tax base? While this analysis does not pretend to have all the answers, it presents straightforward and practical solutions to many of the problems plaguing the nation's transfer tax system using sales to grantor trusts as a case study.

I. INTRODUCTION 974
II. SALES TO GRANTOR TRUSTS: HOW THEY OPERATE AND COMPARE TO OTHER TRANSFER TAX MINIMIZATION TECHNIQUES 975
A. How a Sale to a Grantor Trust Operates 976
B. Comparison to Other Estate Planning Minimization Techniques 982
1. Zeroed-Out GRATs 983
2. Valuation Discounts 985
III. WHAT THE IRS CAN DO TO MAINTAIN THE INTEGRITY OF THE TRANSFER TAX SYSTEM 987
A. Revoke and Rewrite Flawed Revenue Rulings 987
B. Craft New Treasury Regulations 991
1. The Current Deference Standard and Its Application 991
2. The Latitude Afforded the IRS in Responding to Transfer Tax Avoidance Techniques 995
C. The Economic Substance Doctrine and Transfer Taxation 999
IV. WHAT CONGRESS CAN DO TO MAINTAIN THE INTEGRITY OF THE TRANSFER TAX SYSTEM 1004
A. Eliminating Absurdities from the Transfer Tax System 1004
1. The Grantor Trust Rules 1005
2. The Annual Exclusion for Present Interest Gifts 1006
3. The Generation-Skipping Transfer Tax Exemption 1008
B. Act with Alacrity to Remedy Legislative Shortcomings and Oversights 1010
V. CONCLUSION 1012

---

* Soled is a professor of Accounting and Information Systems and Director of the Master of Accounting in Taxation Program at the Rutgers University Business School.

** Gans is a professor of law at Hofstra University, where he is a leading national scholar in the field of gift and estate taxes.

I. INTRODUCTION

In the Tax Relief, Unemployment Insurance Reauthorization, and Job Creation Act of 2010, Congress temporarily raised the applicable exclusion amount—the dollar figure that taxpayers can pass free of transfer tax (i.e., estate, gift, and generation-skipping transfer taxes)—from $1 million to $5 million.[1] This law is set to expire at the end of 2012, at which time the $1 million applicable exclusion amount is scheduled to return.[2] In two years, if Congress wishes to maintain the $5 million applicable exclusion amount and avoid costing the federal coffer billions of dollars in lost revenue, it will have to eliminate several of the most utilized tax-saving devices in estate planning.[3] The devices currently under consideration for elimination include the so-called zeroed-out, grantor-retained annuity trusts (GRATs),[4] minority and marketability valuation discounts for certain intrafamily transfers,[5] and qualified personal residence trusts.[6] These staples of the estate planning world have been part of the panoply of tools that practitioners have devised to minimize taxpayers' transfer tax burdens.

Notwithstanding congressional attention to the elimination of these mainstay planning tools, there has been no discussion in Washington, D.C. to date about eradicating other commonplace transfer tax-savings devices. One such device is known as a sale to a grantor trust, which can replace many of the devices under consideration for the congressional knife and

---

1. Pub. L. No. 111-312, § 303, 124 Stat. 3296 (2010) [hereinafter 2010 Tax Relief Act].

2. *Id.* §§ 101–03.

3. For a discussion pertaining to the possible elimination of several estate planning devices, *see* GENERAL EXPLANATIONS OF THE ADMINISTRATION'S FISCAL YEAR 2011 REVENUE PROPOSALS (Feb. 2010), http://www.treasury.gov/resource-center/tax-policy/Documents/greenbk10.pdf.

4. *See, e.g.*, James M. Delaney, *Split Interest Valuation: The Devil Is in the Detail*, 37 CAP. U. L. REV. 929, 954 (2009) ("With the evolution of the zeroed-out GRAT, the estate planning profession seems to have once again frustrated the goals of the Treasury.").

5. *See, e.g.*, Brant J. Hellwig, *On Discounted Partnership Interests and Adequate Consideration*, 28 VA. TAX REV. 531, 533 (2009) ("Family limited partnerships have dominated the judicial landscape in the estate and gift tax arena for nearly a decade. . . . Their principal advantage lies in the prospect of significant estate and gift tax savings generated through the exploitation of discounts used to value equity interests in closely held entities."). *See generally* Laura E. Cunningham, *Remember the Alamo: The IRS Needs Ammunition in Its Fight Against the FLP*, 86 TAX NOTES 1461 (2000) (describing the legislative action needed to eliminate the tax advantages of the family limited partnership); Leo L. Schmolka, *FLPs and GRATS: What to Do?*, 86 TAX NOTES 1473 (2000) (proposing solutions to several tax loopholes including family limited partnerships).

6. *See, e.g.*, Denver S. Gilliand, *Fractional Interests Make a Better QPRT*, 32 REAL PROP. PROB. & TR. J. 145, 180 (1997) ("[Qualified Personal Residence Trusts], as an exception to the Chapter 14 valuation rules, offer some significant estate tax planning opportunities.").

achieve similar transfer tax savings.[7] While nothing is certain, estate planners will likely switch gears in the aftermath of the likely transfer tax system overhaul and use sales to grantor trusts, among other techniques, to fill the void left by the absence of comparable transfer tax-savings devices.[8]

In anticipation of taxpayers' attempts to minimize their transfer tax obligations, this analysis uses sales to grantor trusts as a case study of what can be done to protect the transfer tax base from erosion. In the sections that follow, we outline how the IRS and Congress should each respond to the emergence of sales to grantor trusts and other transfer tax-savings devices that ultimately become taxpayers' methods of choice to defeat their transfer tax obligations. In Section II, we overview how sales to grantor trusts operate and how they compare to other transfer tax savings devices. In Section III, we point out how such sales and other transfer tax savings devices are vulnerable to challenges by the IRS. In Section IV, we suggest ways that Congress can stem taxpayers' use of sales to grantor trusts and other planning devices designed to circumvent transfer tax obligations. In Section V, we offer our conclusions.

II. SALES TO GRANTOR TRUSTS: HOW THEY OPERATE AND COMPARE TO OTHER TRANSFER TAX MINIMIZATION TECHNIQUES

Close to a century ago, Congress instituted the estate tax;[9] ever since then, taxpayers have sought creative ways to minimize their transfer tax burdens.[10] In the estate planning sphere, some forms of taxpayers' "creativeness" have been deemed impermissible by the courts;[11] however, other such forms have been sanctioned by the courts, and, as a result, they have been added to practitioners' stores of acceptable estate planning

---

7. *See* Robert T. Danforth, *A Proposal for Integrating the Income and Transfer Taxation of Trusts*, 18 VA. TAX REV. 545, 619 (1999) (discussing "the use of grantor trust status as a means of avoiding estate and gift taxes").

8. *Compare* George Cooper, *A Voluntary Tax? New Perspectives on Sophisticated Estate Tax Avoidance*, 77 COLUM. L. REV. 161 (1977) (describing how, when it comes to transfer taxes, taxpayers have devised numerous ways to skirt their obligations), *with* Paul L. Caron & James R. Repetti, *The Estate Tax Non-Gap: Why Repeal a Voluntary Tax?*, 20 STAN. L. & POL'Y REV. 153 (2009) (describing how the estate tax imposes a significant transfer tax burden on most of the nation's largest estates).

9. Act of September 8, 1916, Pub. L. No. 64-271, § 2(b), 39 Stat. 756.

10. *See, e.g.*, Richard Schmalbeck, *Avoiding Federal Wealth Transfer Taxes*, *in* RETHINKING ESTATE AND GIFT TAXATION 113 (William G. Gale, James R. Hines, Jr. & Joel Slemrod eds., 2001) (exploring some of the historical ways in which taxpayers have sought to alleviate their transfer tax burdens).

11. *See, e.g.*, Heyen v. United States, 945 F.2d 359, 365 (10th Cir. 1991) (disregarding as a sham a taxpayer's use of twenty-seven unrelated straw people to obtain twenty-seven additional annual exclusions for gifts to the taxpayer's family).

devices.[12] To date, insofar as courts have not challenged the viability of sales to grantor trusts, such transactions fall squarely within the scope of the latter category. In the subsections below, (A) we explore how a sale to a grantor trust operates, and (B) we compare this technique to other transfer tax minimization techniques.

A. How a Sale to a Grantor Trust Operates

Utilized as a device to achieve transfer tax savings, a sale to grantor trust constitutes a complex arrangement. As set forth below, we outline how practitioners orchestrate this arrangement, its income tax and transfer tax implications, and why taxpayers have found its use attractive from a transfer tax perspective.

Before analyzing this complex arrangement, however, some basic fundamentals are in order. Subchapter J of the Code governs the income taxation of trusts and estates. Subpart E of Subchapter J sets forth special rules for grantor trusts and the fact that such trusts are essentially ignored for income tax purposes (i.e., in most instances, the grantor and the trust are treated as one-and-the-same taxpayer).[13] I.R.C. §§ 673 through 679 set forth the criteria that result in part or all of a trust having grantor trust status.[14] If a trust has grantor trust status, the tax-reporting obligations of such a trust are generally negligible.[15]

Although the separate existence of grantor trusts is generally ignored for income tax purposes, the same fate does not hold true for estate tax purposes. To the contrary, the assets in a grantor trust will not be included in the grantor's gross estate[16] if properly drafted.[17] This disparate tax treatment between income and estate taxes, in which the former ignores

---

12. *See, e.g.*, Patrick T. Neil, *"Bare"ly Legal: The Evolution of Naked Crummey Powers and a Call for Reform*, 30 FLA. ST. U. L. REV. 923 (2003) (discussing how the Tax Court's decision in *Estate of Cristofani v. Comm'r*, 97 T.C. 74 (1991) sanctioned transfers made into a trust for the benefit of remote beneficiaries and how such transfers qualified for the present interest annual exclusion).

13. Treas. Reg. §§ 1.671-1(a) (as amended in 1980), 1.671-3(a)(1) (as amended in 1969); Rev. Rul. 57-390, 1957-2 C.B. 326. *Compare* Rothstein v. United States, 735 F.2d 704, 710 (2d Cir. 1984) (noting that for income tax purposes, a transaction between a taxpayer and a grantor trust should be respected and taxed accordingly), *with* Rev. Rul. 85-13, 1985-1 C.B. 184 (ruling that transactions between a grantor and a grantor trust cannot have income tax significance).

14. *See* Leo L. Schmolka, *Selected Aspects of the Grantor Trust Rules*, *in* THE NINTH ANNUAL INSTITUTE ON ESTATE PLANNING 1400 (1975) (discussing the criteria for treatment of a trust as a grantor trust).

15. *See* Treas. Reg. § 1.671-4 (as amended in 2006) (stating the tax obligations of grantor trusts).

16. *See* I.R.C. § 2033 (2006) (including in a decedent's estate those assets in which decedent held an interest).

17. Danforth, *supra* note 7, at 557.

grantor trusts' separate existence and the latter respects their separate existence, sets the stage for sales to grantor trusts.

Consider how a sale to a grantor trust functions. Suppose Kay owns title to a piece of highly appreciating rental real estate with a current fair market value of $1 million and an adjusted tax basis of $200,000. Kay visits her local estate planning attorney, who renders the following advice: Kay should establish an irrevocable trust, the terms of which provide her with sufficient indicia of control that, for income tax purposes, make it a grantor trust; however, for transfer tax purposes, such indicia of control fall short of causing inclusion of the trust's assets in Kay's gross estate.[18] Once Kay establishes the irrevocable trust with the precise terms described above, Kay's attorney advises her to make a $100,000 cash contribution to the trust. Finally, Kay's attorney proposes that after a sufficient period of time after funding the trust, the trustee of the irrevocable trust, Bea, should purchase title to Kay's appreciating real estate using the $100,000 cash as a down payment and issuing a nine-year promissory note (with $100,000 annual principal payments plus applicable interest) to pay off the balance due.[19] The entirety of this proposed transaction is represented by two simple diagrams as follows:[20]

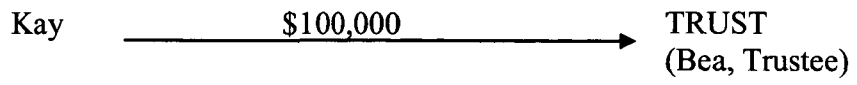
---

18. There are several ways to accomplish this goal. Probably the most common method is to employ I.R.C. § 675(4)(C), which provides that if a person, in a non-fiduciary capacity, has the power to switch title to assets in her own name with assets of equivalent value held by the trust, grantor trust status is appropriate. I.R.C. § 675(4)(C) (2006).

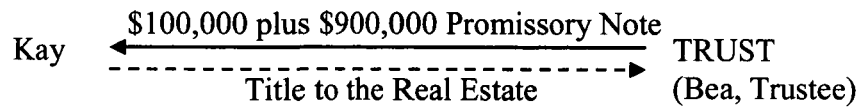
19. When a sale is made to a trust, the IRS may attempt to invoke I.R.C. § 2036 to bring the date-of-death value of the assets sold to the trust into the grantor's gross estate. Even if the grantor does not retain an interest in the trust or the right to control the management of the trust—the two predicates for invoking § 2036—the IRS may be able to sustain this argument. In essence, the argument would be based on the view that, in substance, the grantor's retained right to receive payments under the note constitutes a retained income stream. Thus, if death should occur before the note is fully paid, inclusion in the grantor's gross estate via § 2036 could occur. *See generally* I.R.C. § 2036 (2006) (providing in effect that, as a general matter, the section does not apply if the grantor's retained access ends before death).

If the IRS were to make this argument, the taxpayer would have two possible responses. First, the taxpayer could argue that the retention of the right to receive payments under a note generated by a sale does not constitute the retention of a right within the meaning of § 2036. *See* Fidelity-Phila. Trust Co. v. Smith, 356 U.S. 274, 280 n.8 (1958) (indicating in dicta that in the case of a sale, if payments need not necessarily derive from the property transferred and are not correlated to the income generated, the grantor should not be treated as having retained access); *see also* Rev. Rul. 77-193, 1977-1 C.B. 273 (applying the *Fidelity-Philadelphia* dicta, a case involving an annuity, to an installment sale). In the minds of many practitioners, one approach to satisfying the conditions set forth in *Fidelity-Philadelphia* is if the trust is first funded with sufficient money—often referred to as seed money. *See* Becklenberg v. Comm'r, 273 F.2d 297, 301-02 (7th Cir. 1959) (applying and upholding the ruling in *Fidelity-Philadelphia*). The difficulty with the seed money approach is that the amount of seed money necessary to satisfy the requirements of the dicta is not

STEP 1: Contribution to the Grantor Trust



STEP 2: Trust Purchases the Real Estate Using the Contributed Cash and a Promissory Note



clear. While many suggest that ten percent of the sales prices is sufficient, *see, e.g.*, Michael D. Mulligan, *Sale to a Defective Grantor Trust: An Alternative to a GRAT*, 23 EST. PLAN. 3, 8 (1996), there is no published authority to this effect. The other difficulty with the seed money approach is that in order to provide the trust with seed money, the grantor must make a taxable gift into a trust, and the size of this gift might give rise to the payment of gift tax. With the passage of the 2010 Tax Relief Act, *supra* note 1, this possible tax friction may no longer be a serious impediment insofar as every taxpayer now enjoys the equivalent of a $5 million gift tax exemption, with married couples able to contribute $10 million free of gift tax. I.R.C. § 2505(a) (2006); I.R.C. § 2010(c) (2006). Another approach used by practitioners to satisfy the *Fidelity-Philadelphia* dicta is a guarantee. Under this approach, the beneficiary of the trust guarantees that the note due to the grantor will be paid even if the trust is unable to make payments. The cases cited by the Supreme Court in *Fidelity-Philadelphia* suggest the viability of this approach. The "guarantee approach," however, raises its own issues. For example, if the guarantor does not have sufficient assets, the guarantee may be seen as more of a façade than reality. *See* Estate of Fabric v. Comm'r, 83 T.C. 932 (1984) (finding that exclusion of the transferred assets from the gross estate was proper). Also, questions have been raised as to whether a fee must be charged for the guarantee and, if so, how much. Indeed, at one point, the IRS had suggested that in the absence of a fee, the guarantor should be treated as having made a taxable gift. *See* I.R.S. Priv. Ltr. Rul. 91-13-009 (Dec. 21, 1990) (holding that gift guarantees are considered taxable gifts).

A second possible defense to an IRS challenge is to rely upon the bona fide and full consideration exception to § 2036. Practitioners who counsel about this kind of sale are understandably cautious about relying on the bona fide exception, because (i) if the IRS is able to establish after the death of the grantor that the price was inadequate—even if minimally inadequate—the exception may not be available, and (ii) the IRS may argue that, even if the price was adequate, the tax-driven nature of the transaction renders the bona fide exception inapplicable. Strangi v. Comm'r, 417 F.3d 468 (5th Cir. 2005) (holding that assets were properly included in taxable estate); Estate of Hughes v. Comm'r, 90 T.C.M. (CCH) 630, 635 (2005) (holding that assets were not includable in gross estate because they lacked value).

20. In terms of practical reality, this proposed transaction may engender further complexities as most practitioners would probably recommend that Kay first transfer title to her real estate to a limited liability company, let some time expire, and then only sell a minority portion of her limited liability company membership interest to the trust. By "wrapping" title to her real estate in a limited liability company, Kay will likely command useful valuation discounts. *See infra* Section II.B.2. For heuristic reasons, we have purposefully chosen to ignore the additional complexity entailed by this sort of planning.

Kay's sale of her real estate title to a grantor trust engenders both income tax and transfer tax implications. Recall that the Code ignores the separate existence of a grantor trust for income tax purposes;[21] in a practical sense, what this means is that the Code treats the trust as the grantor's alter ego.[22] The IRS has long ruled that for income tax purposes, transactions that taxpayers engage in with themselves are not recognized.[23] That being the case, when Bea, in her fiduciary capacity as trustee of the grantor trust, purchases the real estate from Kay, Kay recognizes no gain or loss (i.e., Kay and Bea are deemed to be one and the same).[24] Consistent with the nonrecognition concept is that the trust would hold title to the purchased real estate with a carryover basis of $200,000.[25] Due to the grantor trust status of the trust, the rental income earned by the trust during each year of its existence would be reportable on Kay's individual income tax return.[26] However, as installment note and interest payments are made to Kay, Kay would not incur any income tax liability, because the receipt of these payments would be ignored for income tax purposes.[27]

For transfer tax purposes, assuming that the trust terms are properly drafted (i.e., they do not provide Kay with any retained indicia of control

---

21. *See supra* note 13 and accompanying text.

22. Craig D. Bell & Julie A. King, *Sweeping Up the Two Percent Floor:* Scott v. United States *and the Deductibility of Investment Advisory Fees*, 38 REAL PROP. PROB. & TR. J. 589, 594 (2003) ("[G]rantor trusts, . . . are historically treated as the alter egos of the grantors[].").

23. *See* Rev. Rul. 85-13, *supra* note 13; *see also* I.R.S. Priv. Ltr. Rul. 2002-47-006 (Nov. 22, 2002) (ruling that when an owner of two separate trusts transfers a life insurance policy from one trust to the other, the transaction will be disregarded for federal income purposes); I.R.S. Priv. Ltr. Rul. 2002-28-019 (Jul. 12, 2002) ("[A] transaction cannot be recognized as a sale for federal income tax purposes if the same person is treated as owning the purported consideration both before and after the transaction.").

24. *See* Rev. Rul. 85-13, *supra* note 13.

25. *See id.*

26. *See* Jonathan G. Blattmachr & Bridget J. Crawford, *Grantor Trusts and Income Tax Reporting Requirements: A Primer*, 16 PROB. & PROP. 18, 18 (2002) ([R]etention by a trust's grantor, or another person, of certain powers over trust property will cause that grantor (or other person) to be deemed to be the owner for income tax purposes of some or all of the trust property."); John B. Huffaker et al., *Is Income Tax Payment by Grantor-Owner of a Subpart E Trust a Taxable Gift?*, 82 J. TAX'N 202, 203 (1995) (indicating that the grantor of a grantor trust is the owner of the trust for federal income tax purposes even though he may receive none of the benefits of the trust); Jerry Kasner, *Defective IRS Reasoning on Gift Tax Consequences of a Defective Trust*, 66 TAX NOTES 1171, 1171 (1995) ("The income of the trust is taxed to the grantor, who is treated as the 'owner' of the trust for federal income tax purposes.").

27. *See* Rev. Rul. 85-13, *supra* note 13.

over the trust assets), title to the trust property and accumulated rental income should not be includable in Kay's gross estate at the time of her death.[28] That being the case, even if the real estate's fair market value soared to $5 million, its entire value would escape inclusion in Kay's taxable estate.

The inconsistency between the income tax (which treats the trust as the grantor's alter ego) and the estate tax (which treats the trust as a stand-alone entity, separate from the grantor) is glaringly apparent and has long existed. Notwithstanding entreaties from academics and other commentators to harmonize the differences between the income and estate taxes,[29] this disparate tax treatment enables taxpayers to reap rich transfer tax savings.[30]

The benefits of this disparate tax treatment are essentially twofold. First, during the trust term, taxpayers engaging in this technique are obligated to pay the trust's income tax.[31] While the benefit of these payments inures to the trust beneficiaries, such payments are not considered transfers that are subject to gift tax.[32] To illustrate, suppose in our previous example that the trust annually earns $100,000 of rental income and that the effective income tax rate is forty-five percent. In the absence of the grantor trust rules, the trust bears the tax burden, resulting in the trust retaining only $55,000 ($100,000 of income less $45,000 of taxes ($100,000 × 0.45)). But the grantor trust nature of the trust, instead, obligates the grantor to pay the tax on the rental income,[33] resulting in the $100,000 of rental income remaining intact and inuring to the benefit of the trust beneficiaries.

A second feature of this sales technique is that, for transfer tax purposes, it enables taxpayers to "freeze" the value of the transferred property.[34] To illustrate, return once again to our example in which Kay anticipated that the real estate in question would appreciate greatly in value. If it appreciated in her name, the initial fair market value plus its appreciation would be includable in Kay's gross estate. Instead, by making this sale, Kay will ultimately receive $1 million ($100,000 cash down

---

28. *See* I.R.C. §§ 2036–38 (2006).

29. *See, e.g.*, Danforth, *supra* note 7.

30. *See, e.g.*, Ronald D. Aucutt, *Installment Sales to Grantor Trusts*, SR034 ALI-ABA 1013 (2010) ("[A sale to grantor trust] is in effect an estate freeze technique that capitalizes on the lack of symmetry between the income tax rules governing grantor trusts and the estate tax rules governing includibility in the gross estate.").

31. I.R.C. § 671 (2006); *see also* Rev. Rul. 04-64, 2004-2 C.B. 7 (ruling that a settlor's payment of income taxes attributable to a grantor trust is not a taxable gift).

32. Rev. Rul. 04-64.

33. *See id.*

34. *See, e.g.*, Karen Burke, *Valuation Freezes After the 1988 Act: The Impact of 2036(c) on Closely-Held Businesses*, 31 WM. & MARY L. REV. 67, 70 (1989) ("In its simplest form, an estate freeze involves the transfer of an interest representing future appreciation by an older generation transferor, coupled with the retention of another interest having a fixed value.").

payment and a promissory note worth $900,000). This $1 million figure represents a freeze of the amount includable in Kay's taxable estate.

The ultimate transfer tax savings associated with such sales can be intoxicating to taxpayers. Compare the overall tax consequences were Kay to (1) retain title to the real estate versus (2) sell this property to a grantor trust, using the following set of assumptions: Kay's real estate generates $100,000 of rental income annually; its value gradually increases to $5 million; the estate and income tax rates are each a flat forty-five percent; and Kay dies in Year 10 immediately after the promissory note has been paid in full.[35]

Retention of the Real Estate: If Kay retained title to the real estate, $5,550,000 would have been includable in her taxable estate (i.e., title to real estate worth $5,000,000 plus the after-tax rental income of $55,000 that the property annually generated ($100,000 – $45,000 ($100,000 × 0.45)) over a ten-year period. The estate tax on $5,550,000 would be $2,497,500 ($5,550,000 × 0.45), leaving a net amount of $3,052,500 to Kay's family.

Sale to Grantor Trust: Suppose instead that Kay sold title to the foregoing real estate to a grantor trust. In this case, only $1,000,000 would be includable in Kay's taxable estate (i.e., the initial $100,000 cash down payment plus the aggregate $900,000 principal payment),[36] producing an overall estate tax liability of $450,000. Under this scenario, Kay's family would own title to $5,000,000 of real estate, $1 million cash from the rental income (free of income tax because Kay, under the grantor trust rules, bore this burden), plus the after-transfer tax cash bequest of $550,000 (i.e., Kay's estate has $1,000,000 from the aggregate installment payments less the presumed $450,000 transfer tax obligation ($1 million times the forty-five percent assumed transfer tax rate)), leaving a net amount of $6,550,000 to Kay's family.

The significant wealth outcome variations produced under scenarios (1) and (2) illustrate a driving force behind taxpayers' motivations to use sales to grantor trusts.[37]

---

35. For purposes of this illustration, assume that Kay had other assets in her estate that absorbed her entire applicable exclusion amount via I.R.C. § 2010.

36. For purposes of this problem, we assumed that Kay used the interest payments made on the promissory note to meet her annual income tax obligation arising with respect to the $100,000 rental income that the trust generated. For example, if the promissory note's interest rate was ten percent, in addition to making its first installment payment of $100,000, the trust would also pay Kay $90,000 of interest ($900,000 x 10%). Kay could use this interest income to meet her $45,000 income tax obligation resulting from the grantor trust nature of the trust. In later years, when the principal balance of the promissory note is much smaller, the interest payments due to Kay will be correspondingly smaller as well (e.g., in Year 9, the trust's final payment would be $110,000, consisting of $100,000 of principal and $10,000 of interest). In later years, the excess interest payments from earlier years could be used to meet the $45,000 income tax obligation on the trust's annual rental income.

37. In the context of a sale to grantor trust scenario, Kay's beneficiaries will have a

B. Comparison to Other Estate Planning Minimization Techniques

In the past, in lieu of sales to grantor trusts, taxpayers sometimes used other estate planning techniques to minimize their transfer tax burdens.[38] The reason for taxpayers' reluctance to employ this technique centers upon a concern about inadvertently triggering gift tax liability should the value of the sold assets be determined to be in excess of the sales price. (This concern, however, will likely dissipate with the increase in the gift tax exemption of $5 million adopted in 2010.)

As a general proposition, many taxpayers lack familiarity with how our tax system operates, and this lack of familiarity is particularly acute in the area of transfer taxes. Given this background, imagine the confusion that discussions of sales to grantor trusts must engender. For starters, tax professionals explaining how this technique operates will often refer to the purchasing trust as being "defective" for income tax purposes.[39] Why? This is because the terms of the trust are designed to achieve grantor trust status, and for the majority of the time that the income tax has been in existence, this status was steadfastly avoided.[40] This avoidance is due to the fact that the income of non-grantor trusts historically has been taxed at lower marginal tax rates than that of individual taxpayers, and the income of all grantor trusts was generally taxed at the grantor's higher marginal tax rates.[41] Even though the Tax Reform Act of 1986 eradicated the tax bracket

---

lower tax basis in the real estate. More specifically, had Kay held title to the property until her death, her tax basis in the real estate would have equaled $5 million in her beneficiaries' hands. *See* I.R.C. § 1014(a) (2006). By contrast, by engaging in this sale to grantor trust, Kay's beneficiaries would ultimately have a $200,000 tax basis in the transferred real estate (i.e., Kay's initial cost basis). This tax basis differential can ultimately result in Kay's beneficiaries bearing a larger income tax burden if and when they were to sell this real estate. *See* I.R.C. § 1001(a) (2006). Note, however, that the law regarding the beneficiaries' basis is not entirely clear. *See* Jonathan G. Blattmachr, Mitchell M. Gans & Hugh H. Jacobson, *Income Tax Effects of Termination of Grantor Trust Status by Reason of the Grantor's Death*, 97 J. TAX'N 149, 158–59 (2002) (suggesting that the tax basis in the beneficiaries' hands might, indeed, equal the sales price). *But see* I.R.S. C.C.A. 200937028 (Sept. 11, 2009) (rejecting this position).

38. *See supra* notes 4–6.

39. *See, e.g.*, Stephen R. Akers, Jonathan G. Blattmachr & F. Ladson Boyle, *Creating Intentional Grantor Trusts*, 44 REAL PROP. TR. & EST. L.J. 207, 211 (2009) ("The term defective was applied first to grantor trusts when the grantor trust rules originally were adopted because, as a general matter, a grantor trust classification prevented income splitting. Avoiding grantor trust status was the typical taxpayer goal. Thus, before 1987 the trust was "defective" from the perspective that the trust income was taxable to the grantor instead of the trust or a trust beneficiary. That label has carried over to today, although now grantor trust status usually is viewed as beneficial.").

40. *Id.*

41. *See generally* Mark L. Ascher, *The Grantor Trust Rules Should Be Repealed*, 96 IOWA L. REV. 885 (2011).

advantage afforded to non-grantor trusts,[42] the putative scourge of being a grantor trust has remained; hence, the moniker defective remains extant. Once a tax professional gets beyond explaining the fact that the trust is not truly defective, the professional then must explain the numerous steps entailed in arranging such trust sales.

Aside from taxpayers' tendencies to shy away from those tax-saving techniques that they cannot readily comprehend and that have a facade of artificiality, taxpayers have discovered recently the appeal of other techniques. The appeal of other techniques is commonly twofold: they enjoy the imprimatur of the IRS, Congress, and/or the courts; and they can produce more bountiful transfer tax savings relative to those offered by sales to grantor trusts.[43] Two such techniques are: (1) zeroed-out GRATs and (2) valuation discounts.

1. Zeroed-Out GRATs

The framework for a zeroed-out GRAT is found in I.R.C. § 2702, entitled "Special valuation rules in case of transfers of interest in trusts."[44] The statute provides that "[t]he value of any retained interest which is not a qualified interest shall be treated as being zero."[45] For example, under this rule, if a taxpayer contributes $1 million into a trust established for the benefit of his children and retains a ten-year income interest (i.e., a nonqualified interest under the Code), the value of the retained income interest is deemed to be zero.[46] Accordingly, notwithstanding the taxpayer's retention of this valuable income right, the taxpayer will nevertheless be deemed to have made a taxable gift of the entire $1 million trust contribution.[47] Congress devised this approach in order to eliminate an abusive strategy under which wealth could be moved to children and others without the full payment of gift tax.[48]

---

42. *See* I.R.C. § 101(a) (2006).

43. *See, e.g.*, Ellen K. Harrison, *A Comparison of Retained Annuities and Sales to Grantor Trusts*, WL SD10 ALI-ABA 763 (1998) ("In most cases, a GRAT may be structured to produce a minimal gift tax value . . . .").

44. I.R.C. § 2702(a)(2)(A) (2006).

45. *Id.*

46. *See id.*

47. *See* I.R.C. § 2702(a)(1).

48. In the years leading up to the enactment of I.R.C. § 2702, taxpayers established grantor-retained income trusts (commonly known in the estate planning community as "GRITs"). In computing the value of a taxpayer's retained-income interest, the Code provided a rate of return equal to the I.R.C. § 7520 rate. This rate of return often made the taxpayer's retained interest appear robust and, in contradiction, the amount of the remainder interest (i.e., the gift) small. Meanwhile, the GRIT trustee could invest in growth assets that produced very little income; by engaging in this kind of investment strategy, the investment growth would inure to the trust beneficiaries and essentially escape any transfer tax

The statute provides a major exception to the general rule of treating the value of any retained interest as zero.[49] It does so by defining a "qualified interest" as including "any interest which consists of the right to receive fixed amounts payable not less frequently than annually."[50] By defining a qualified interest in this straightforward fashion, Congress thought it had eliminated opportunities for valuation abuse. Retained amounts that are "fixed" appear to be safe from taxpayer manipulation, and Congress affixed a rate of return on the contributed trust property equal to the I.R.C. § 7520 rate (a rate that is issued on a monthly basis and is equal to 120 percent of the federal midterm interest rate).[51] The opportunity for taxpayers to game an arrangement of this sort seemed remote, because the value of the remainder interest could be determined with apparent accuracy.[52]

Working within statutory parameters, crafty taxpayers instead designed GRATs to be the perfect transfer tax loophole. Taxpayers' strategy was simple: Contribute property that produced a rate of return in excess of the § 7520 rate and retain robust annuity payments such that the value of the retained interest essentially equaled the value of the remainder interest (e.g., contribute $1 million into a trust but retain an interest therein slightly less than or equal to $1 million, producing little or no taxable gift).[53] If the trust property produced a rate of return that was in excess of the § 7520 rate, assets would remain in the trust and could pass tax-free to the named trust beneficiaries.[54] If the trust assets failed to produce the § 7520 rate of return, nothing would be left in the trust, but because the up-front gift was deemed to be zero or de minimis, the taxpayer suffered no negative consequences for making this trust contribution.[55] This technique, known as a "zeroed-out GRAT," has become an essentially foolproof method of achieving transfer tax savings with no downside risk and even the Tax Court has tacitly sanctioned its use.[56]

---

exposure. *See generally* Mitchell Gans, *GRIT's, GRAT's, and GRUT's: Planning and Policy*, 11 VA. TAX REV. 761 (1992) (explaining how taxpayers would exploit the use of GRITs in fashions designed to minimize their transfer tax burdens).

49. *See* I.R.C. § 2702(a)(2)(A) (2006).

50. I.R.C. § 2702(b)(1).

51. I.R.C. § 7520(a)(2) (2006).

52. *See* Grayson M. P. McCouch, *Rethinking Section 2702*, 2 FLA. TAX REV. 99, 99 (1994) ("In 1990, Congress added chapter 14 to the Code to address several gift and estate tax avoidance techniques that flourished under prior law." (footnotes omitted)).

53. KATHRYN G. HENKEL, ESTATE PLANNING AND WEALTH PRESERVATION: STRATEGIES AND SOLUTIONS ¶ 22.03[2][b] (1997).

54. *See* I.R.C. §§ 2702 (a)(2)(B), 7520(a).

55. *See* I.R.C. § 2702(a)(2)(A).

56. *See* Walton v. Comm'r, 115 T.C. 589 (2000), *acq.*, IRS Notice 2003-72, 2003-2 C.B. 964 (ruling that for purposes of I.R.C. § 2702, if the transferor were to die during the term of the retained interest, annuity payments that were to continue to be made to the

Consider the following example. In September 2011, when the § 7520 rate was two percent, a taxpayer contributed $1 million into a two-year GRAT and retained an annual annuity payment of $520,562. As a result of retaining such a large annuity interest, the value of the retained interest was deemed equal to $1 million, making the value of the corresponding taxable gift equal to zero ($1 million trust contribution less $1 million retained interest). At the end of Year 1, the taxpayer will receive an annuity payment of $520,562, and after Year 2, the taxpayer will receive another annuity payment of $520,562. At the end of the two-year period after making the two annuity payments, if nothing remains in the trust, the GRAT will prove unsuccessful. Conversely, at the end of the two-year period after making the two annuity payments, if something remains in the trust, the GRAT will prove successful, and whatever remains in the trust will transfer tax-free to the trust's named beneficiaries. Given the absence of a downside risk, it comes as no surprise that the establishment of zeroed-out GRATs has spiraled as their use is regularly promoted by estate planners.

2. Valuation Discounts

In addition to zeroed-out GRATs, another common estate planning technique is the use of valuation discounting. Such discounting can produce stellar transfer tax results.[57] Under this methodology, taxpayers typically gift or sell interests in closely held businesses. Consider the salient fact that the interests in such business enterprises are not publicly traded on a recognized exchange, and the owner of a non-controlling interest is unable to direct the management of the entity. For transfer tax purposes, the absence of a ready market coupled with a lack of control typically reduces the value of such closely held business interests—often producing so-called minority and marketability valuation discounts of thirty percent or more[58] under the traditional "willing buyer/willing seller" test.[59]

---

transferor's estate constituted a qualified interest, thereby reducing the value of the remainder interest). Note, however, that the IRS insists that the preamble to the regulations under § 2702 does not contemplate that a GRAT can be zeroed out. *See* Tech. Adv. Memo. 2002-45-053 (Nov. 8, 2002).

57. *See generally* Brant J. Hellwig, *On Discounted Partnership Interests and Adequate Consideration*, 28 VA. TAX REV. 531 (2009); Mitchell M. Gans, *Deference and Family Limited Partnerships: A Case Study*, 39 INST. ON EST. PLAN. ¶ 500 (Tina Portuondo ed., 2005).

58. *See* Louis A. Mezzullo, *Valuation of Corporate Stock*, 831-3d TAX MGM'T PORTFOLIO worksheet 1 (2007); *see also* Brant J. Hellwig, *Revisiting* Bryum, 23 VA. TAX REV. 275, 278–79 (2003) ("With courts frequently sustaining combined minority-interest and marketability discounts in the range of 30–50% from proportionate value, the use of limited partnerships for estate-planning purposes is widely regarded as undermining the integrity of the estate tax." (footnotes omitted)).

59. *See* Treas. Reg. § 20.2031-1(b) (1965) ("The fair market value is the price at

Consider the following example. A taxpayer has a $1 million piece of rental real estate. She contributes this real estate to a limited liability company and then gifts a twenty-five percent membership interest in this limited liability company to each of her four children. Each membership interest is not traded on a public exchange and represents a minority interest in this enterprise; thus, rather than reporting a $1 million taxable gift for gift tax reporting purposes, the taxpayer instead can likely show a taxable gift equal to $600,000 ($1,000,000 less $400,000 ($1,000,000 times forty percent discount attributable to the nature of the transferred membership interests)).[60] Essentially, by wrapping this property in the form of a limited liability company, the taxpayer is able to make $400,000 of value disappear from the transfer tax base.[61] Although this sleight of hand seems too good to be true, over a decade ago in a major concession to taxpayers, the IRS gave this technique its imprimatur of approval.[62]

Properly structured zeroed-out GRATs and the use of valuation discounts illustrate that when it comes to transfer tax minimization techniques, taxpayers have had a myriad of options. What made these transfer tax savings options particularly attractive is that they presented taxpayers with little downside risk, and they were readily comprehensible.[63] As a result, for the past two decades, taxpayers have continuously exploited these and several other transfer tax savings techniques.[64]

Assuming that Congress will at some point eliminate several of the most utilized estate planning techniques, including zeroed-out GRATs and valuation discounts, resourceful taxpayers will seek alternative means to minimize their transfer tax burdens. Therefore, they will likely turn to sales

---

which the property would change hands between a willing buyer and a willing seller, neither being under any compulsion to buy or to sell and both having reasonable knowledge of relevant facts.").

60. *Id.*

61. Karen C. Burke & Grayson M.P. McCouch, *Family Limited Partnerships: Discounts, Options, and Disappearing Value*, 6 FLA. TAX REV. 649, 650 (2004).

62. Rev. Rul. 93-12, 1993-1 C.B. 202.

63. While the sale of a non-controlling interest may present a downside risk—the IRS could argue that the sales price was inadequate and that a gift tax should therefore be paid—estate planners have devised strategies that minimize or eliminate such risk. *See, e.g.*, Petter v. Comm'r, 98 T.C.M. (CCH) 534 (2009) (rejecting the IRS's attempt to revalue the sold/gifted units based on a clause inserted in the documents that would divert units to charity in the event of such a revaluation).

64. In the estate tax area, a quick review of continuing legal education programs signifies these techniques' prominence. *See generally* Steve R. Akers, *Advanced Transfer Planning, Including Strategies to Maximize Benefits of GRATS and Sales to Grantor Trusts Given Recent Market Declines*, SR002 ALI-ABA 801 (2009); Lawrence P. Katzenstein, *Some Interest-Sensitive Estate Planning Techniques (with an Emphasis on GRATS and QPRTS)*, SR034 ALI-ABA 109 (2010); William D. Kirchick, *Using GRATS in a Down Economy*, SR013 ALI-ABA 211 (2009); David Pratt, *Update on Use of Family Limited Partnerships and Discount Planning*, SP037 ALI-ABA 399 (2009).

to grantor trusts and other tax-saving techniques, enduring the complexity and artificiality that such techniques engender. The IRS and Congress, however, do not have to be wallflowers and allow taxpayers free rein to subvert the transfer tax base. Instead, as discussed in the next two sections, the IRS and Congress have many weapons at their disposal to defeat such sales and other estate planning techniques.

III. WHAT THE IRS CAN DO TO MAINTAIN THE INTEGRITY OF THE TRANSFER TAX SYSTEM

The IRS has several weapons in its arsenal to defeat the kinds of transactions that Congress did not expressly authorize or sanction and that erode the transfer tax base.[65] In the sections that follow, we outline how the IRS is at liberty to (A) revoke and rewrite flawed revenue rulings, (B) craft new Treasury regulations, and (C) invoke the application of the codified economic substance doctrine. Using this latitude, the IRS can eliminate sales to grantor trusts and other techniques that subvert the transfer tax system.

A. Revoke and Rewrite Flawed Revenue Rulings

The IRS promulgates its administrative positions in several different forms, including revenue rulings, revenue procedures, notices, and announcements.[66] Among these forms, the IRS has historically articulated some of its most important positions via revenue rulings. "Revenue rulings are official interpretations by the Service, which are prepared in the Associate Chief Counsel Offices and published in the Internal Revenue Bulletin by the Service."[67] Such rulings represent the IRS's position on a particular set of facts and conclusions of law and are published with the intention that a particular issue will be handled with uniformity throughout the country.[68]

Once the IRS issues a revenue ruling, the agency has the prerogative to change it. And despite the fact that the IRS can exercise this power retroactively,[69] it generally only exercises this power prospectively.[70] Many

65. The transfer tax base is very broad and is theoretically designed to capture all lifetime transfers, *see* I.R.C. § 2512 (2006), and include all property owned directly and indirectly at death, *see* I.R.C. §§ 2031, 2035–38 (2006).

66. *See generally* Donald L. Korb, *The Four R's Revisited: Regulations, Rulings, Reliance, and Retroactivity in the 21st Century: A View from Within*, 46 DUQ. L. REV. 323 (2008) (former chief counsel for the IRS explaining the nature of administrative information that the IRS issues).

67. *Id.* at 330.

68. Treas. Reg. § 601.601(d) (as amended in 1987).

69. *See* Dixon v. United States, 381 U.S. 68, 74–75 (1965) (explaining why the IRS can retroactively apply its own rulings). *But see* Silco, Inc. v. Comm'r, 779 F.2d 282, 286–

times, the impetus underlying such administrative changes is a well-reasoned court ruling that results in an IRS loss;[71] other times, the IRS will revoke a prior ruling if the agency believes its position is simply contrary to the existing state of the law.[72]

When it comes to the tax consequences associated with sales to grantor trusts, the IRS has promulgated Revenue Ruling 85-13, which ironically lays the foundation for the acceptability of such transactions.[73] *Rothstein v. United States*[74] is the case that led the IRS to issue this revenue ruling.[75] In *Rothstein*, the taxpayer had established an irrevocable trust in which the taxpayer's wife was the sole trustee and his children were the income beneficiaries.[76] Several years after establishing the trust, the taxpayer purchased shares of stock in a closely held corporation from the trustee, using as consideration an unsecured promissory note.[77] Subsequent to this exchange (which, according to the taxpayer's position, did not give rise to taxable income), the taxpayer liquidated the closely held corporation.[78] By using the shares' purchase price (i.e., the amount of the note) as the shares' cost basis, the taxpayer then claimed that the corporate liquidation gave rise to a short-term capital loss.[79]

In adjudicating this case, there were two issues that required resolution. The first was whether the terms of the trust were such that it should be classified as a grantor trust, and the second was the tax consequences that flowed from this classification.[80]

---

87 (5th Cir. 1986) (taxpayer's position upheld based upon a reliance on previously issued revenue rulings); Rauenhorst v. Comm'r, 119 T.C. 157, 170, 172 (2002) (finding that, generally, the Tax Court treats a revenue ruling as a concession by the IRS that the agency must either withdraw or modify before it can take a contrary position).

70. *See* Treas. Reg. § 601.601(d)(2)(v)(c) (as amended in 1987) ("Where Revenue Rulings revoke or modify rulings previously published in the Bulletin the authority of I.R.C. § 7805(b) of the Code ordinarily is invoked to provide that the new rulings will not be applied retroactively to the extent that the new rulings have adverse tax consequences to taxpayers.").

71. *See, e.g.*, Estate of Wall v. Comm'r, 101 T.C. 300, 312–13 (1993) (resulting in the IRS's revocation of Rev. Rul. 79-353, 1979-2 C.B. 325, and replacing it with Rev. Rul. 95-53, 1995-2 C.B. 191, the terms of which are consistent with the *Wall* decision).

72. *See, e.g.*, Rev. Rul. 93-12, 1993-1 C.B. 202 (acknowledging, after suffering a series of court defeats over a period of several years, that minor and marketability discounts are permissible for gratuitous transfers of closely held business interests).

73. *See supra* note 13 and accompanying text.

74. Rothstein v. United States, 574 F. Supp. 19 (D. Conn. 1983), *rev'd*, 735 F.2d 704 (2d Cir. 1984).

75. *See id.*; Rev. Rul. 85-13, 1985-1 C.B. 184.

76. *Rothstein*, 574 F. Supp. at 20.

77. *Id.* at 20–21.

78. *Id.* at 21.

79. *Id.*

80. *Id.* at 22.

At the district court level, the IRS prevailed: the court ruled that because the grantor could indirectly borrow from the trust, the trust was a grantor trust.[81] The court further found that the taxpayer had a carryover tax basis in the corporate shares (i.e., equal to the tax basis in the hands of the grantor trust) and disallowed the taxpayer's putative loss on the corporate liquidation.[82] To illustrate this with numbers, if the trust had a $10 basis per share in the corporation in question and the purchase price paid by the grantor was $100 per share, the court would rule that the purchasing taxpayer had a $10 basis per share. If the taxpayer then liquidated the corporation and the liquidation proceeds were $30 per share, the taxpayer would have experienced a $20 gain per share (i.e., $30 – $10).

The U.S. Court of Appeals for the Second Circuit reversed the district court's decision.[83] While it agreed with the lower court's classification analysis (i.e., the instrument in question established a grantor trust), it disagreed with the consequences that stemmed from this classification.[84] Writing for the majority, Judge Friendly read the grantor trust statute in a literal fashion, claiming that it attributed items of income, deductions, and credits from a grantor trust to its grantor.[85] In all other respects, the terms of the Code were fully applicable, including the provisions stating that a sales transaction gives rise to gain or loss,[86] and an asset's purchase price constitutes its cost basis.[87] Accordingly, on the corporate liquidation, the court allowed the taxpayer a loss, because the amount of the liquidation proceeds were less than the taxpayer's cost basis in his shares.[88] For illustration purposes once again, if the trust had a $10 basis per share in the corporation in question and the purchase price was $100 per share, the Second Circuit would rule that the purchasing taxpayer had a $100 basis per share. If the taxpayer then liquidated the corporation and the liquidation proceeds were $30 per share, the taxpayer would experience a $70 loss per share (i.e., $30 – $100).[89]

---

81. *Id.* at 23.

82. *Id.*

83. Rothstein v. United States, 735 F.2d 704, 710 (2d Cir. 1984).

84. *See id.* at 708–09.

85. *See id.* at 709 ("[Section] 671 dictates that, when the grantor is regarded as 'owner,' the trust's income shall be attributed to him—this and nothing more.").

86. I.R.C. § 1001.

87. I.R.C. § 1012 (2006 & Supp. III 2009).

88. *Rothstein*, 735 F.2d at 710.

89. Although the sale of the corporate stock by the trustee to the taxpayer gave rise to a gain, this gain was apparently able to be reported on the installment method, a fact that apparently irritated the government. *See id.* ("The Government's grievance apparently derives from the fact that, . . . neither the taxpayer nor the trust had reported a capital gain on the sale of the IDI shares in 1964."). Congress has since put a limitation on the use of the installment method in this context. I.R.C. § 453(e) (2006). *See, e.g.*, Shelton v. Comm'r, 105 T.C. 114, 120 (1995) (In return for an installment note, a taxpayer sold the stock of a closely

After suffering a defeat at the hands of the Second Circuit, the IRS hastily issued Revenue Ruling 85-13, announcing that it would not follow the *Rothstein* decision, because it recognized the separate existence of a grantor trust for tax reporting purposes.[90] In other words, the agency would continue to ignore all dealings that a taxpayer had with respect to a grantor trust, including the recognition of gains and losses and tax basis determination issues.[91]

In retrospect, the IRS should have applauded the *Rothstein* outcome. The Second Circuit's holding makes clear that transactions between a taxpayer and a grantor trust should not be ignored. While its holding might seem to invite opportunities for abuse (e.g., taxpayers could engage in transactions with their alter egos (grantor trusts)), losses arising between related parties are traditionally disallowed;[92] and, if necessary, Congress could have crafted a provision requiring that gains arising from such related-party transactions could not be used to absorb capital losses or net operating losses.

As long as Revenue Ruling 85-13 is retained, it will serve to sanction the use of sales to grantor trusts. As such, it will enable taxpayers to use such sales as tools to chisel away at their transfer tax obligations. As with any revenue ruling that taxpayers use as a mechanism to defeat their tax obligations, the IRS should examine the merits of the ruling and the risks associated with revoking it. An analysis by the IRS would likely reveal that the agency should revoke Revenue Ruling 85-13—and for that matter, any other revenue ruling that taxpayers use to defeat legitimate tax obligations—and let the *Rothstein* decision stand.[93] Put differently, the IRS should not allow itself to be a wallflower, passively watching as the federal coffers are drained.

---

held business to related parties who subsequently liquidated the company; the Tax Court held that this liquidation was a second disposition and, as such, was subject to I.R.C. § 453(e), which "was enacted as a response to the use by taxpayers of installment sales to a related intermediary in order to defer recognition of gain while at the same time effectively realizing appreciation on the property by means of a resale to a party outside the 'related group' for an immediate payment.").

90. Rev. Rul. 85-13, 1985-1 C.B. 184.

91. *See id.*

92. I.R.C. § 267(a) (2006).

93. Note the fact that a longstanding revenue ruling followed by the issuance of a contrary regulation does not undercut an agency's deference claim. *See, e.g.*, Smiley v. Citibank (S. D.), N.A., 517 U.S. 735, 740-41 (1996). Instead of issuing new regulations, if the IRS simply issued a new revenue ruling, it would only be entitled to limited deference. Skidmore v. Swift & Co., 323 U.S. 134 (1944).

B. Craft New Treasury Regulations

The IRS has a potent weapon at its disposal against which taxpayers in general have a difficult time avoiding. Specifically, in interpreting the Code, the IRS can issue regulations that command deference from the courts. We analyze below (1) the current deference standard and its application and (2) the latitude that deference affords the IRS in responding to transfer tax avoidance techniques such as sales to grantor trusts.

1. The Current Deference Standard and Its Application

The origin of how courts defer to administrative agency decisions is long, difficult, and complex to trace.[94] However, the current deference standard, embodied in *Chevron U.S.A., Inc. v. Natural Resources Defense Counsel, Inc.*,[95] and reinforced by the recent Supreme Court decision in *Mayo Foundation for Medical Education & Research v. United States*,[96] is relatively easy to understand and apply.

By way of background, the IRS is able to issue "legislative" regulations (i.e., regulations that frame a specific body of law) in accordance with a specific delegation from Congress.[97] Under the general authority of I.R.C. § 7805(a), the service can also issue "interpretative" regulations (i.e., regulations that offer guidance as to what a specific statutory body of language means or signifies).[98] The courts had applied different levels of deference to these two types of regulations.[99] In *Mayo Foundation*, the

---

94. *See generally* Mitchell M. Gans, *Deference and the End of Tax Practice*, 36 REAL PROP. PROB. & TR. J. 731 (2002) (describing the history of judicial deference to administrative interpretations).

95. Chevron U.S.A., Inc. v. Natural Res. Def. Counsel, Inc., 467 U.S. 837 (1984).

96. Mayo Found. for Med. Educ. & Research v. United States, 131 S.Ct. 704 (2011).

97. *See* Kristin E. Hickman, *Agency Specific Precedents: Rational Ignorance or Deliberate Strategy?*, 89 TEX. L. REV 89, 104 (2011) (explaining the history of regulations issued pursuant to a general grant of authority and specific grant of authority); *see also* I.R.C. § 1502 (2006) (authorizing the Treasury Department to promulgate regulations that delineate how those corporations that qualify can file tax returns on a consolidated basis); N.Y. STATE BAR ASS'N TAX SECTION, REPORT ON LEGISLATIVE GRANTS OF REGULATORY AUTHORITY 1, 2–6 (Nov. 3, 2006), *available at* http://www.nysba.org/Content/ContentFolders20/TaxLawSection/TaxReports/1121Report.pdf (estimating that there are approximately 550 such provisions in the Code).

98. *See* Tutor-Saliba Corp. v. Comm'r, 115 T.C. 1, 7 (2000) ("An interpretive regulation is issued under the general authority vested in the Secretary [of the Treasury] by section 7805, . . ."); Comment, *Denying Retroactive Effect to Invalidation of Administrative Rules*, 12 STAN. L. REV. 826, 830-31 n.25 (1960) ("Treasury regulations, often classified as interpretive, are regarded as merely stating the Treasury's construction of the statute.").

99. *See Mayo Found.*, 131 S.Ct. at 713; ("In two decisions predating *Chevron*, this Court stated that 'we owe the [Treasury Department's] interpretation less deference' when it is contained in a rule adopted under that 'general authority' than when it is 'issued under a

Supreme Court stated that the same standard of deference, the *Chevron* standard, should apply to both types of regulations.[100]

In determining whether an administrative agency's statutory construction should be upheld, the Supreme Court in *Chevron* enunciated a two-step analysis.[101] Step one is to determine whether "the intent of Congress is clear"; if it is, ". . . the agency[] must give effect to the unambiguous expressed intent of Congress."[102] Step two occurs only if the intent of Congress is unclear (i.e., the statute is silent or ambiguous with respect to a specific issue); in those situations, "the question for the court to determine is whether the agency's answer is based on a permissible construction of the statute."[103] Regarding step two, the Supreme Court, in another case, noted that "we defer to the Commissioner's regulations as long as they 'implement the congressional mandate in some reasonable manner.'"[104] Accordingly, only when an agency's construction of an unclear statute is unreasonable does judicial deference to an agency's rule-making authority end.[105]

Swallows Holding, Ltd. v. Commissioner is an illustrative case that demonstrates how the application of the *Chevron* standard applies to Treasury Department regulations.[106] The facts contained in *Swallows Holding* are straightforward: the taxpayer was a Barbados corporation that owned U.S. rental property.[107] This rental property generated income and experienced concomitant deductions; however, the taxpayer failed to file tax returns for tax years 1993, 1994, 1995, and 1996, delaying such submissions until 1999.[108]

I.R.C. § 882(c)(2) provides that a foreign corporation "shall receive the benefit of the deductions . . . allowed to it . . . only by filing or causing to be filed . . . a true and accurate return, in the manner prescribed in subtitle F . . . ."[109] On the basis of this section, because the taxpayer failed to file timely

---

specific grant of authority to define a statutory term or prescribe a method of executing a statutory provision.'"). *See generally* Gans, *supra* note 94 (describing the history of judicial deference to administrative interpretations).

100. *Mayo Found.*, 131 S.Ct. at 714.

101. *Chevron*, 467 U.S. at 842–43.

102. *Id.*

103. *Id.* at 843.

104. United States v. Cleveland Indians Baseball Co., 532 U.S. 200, 219 (2001) (quoting United States v. Correll, 389 U.S. 299, 307 (1967)).

105. Swallows Holding, Ltd. v. Comm'r, 515 F.3d 162, 169 (3rd Cir. 2008).

106. *See id.* at 162. *See generally* Mark E. Berg, *Judicial Deference to Tax Regulations: A Reconsideration in Light of* National Cable, Swallows Holding, *and Other Developments*, 61 TAX LAW. 481 (2008) (discussing judicial deference in light of recent cases including *Swallows Holding*).

107. *Swallows Holding*, 515 F.3d at 165.

108. *Id.*

109. I.R.C. § 882(c)(2) (2006).

tax returns for the tax years in question, the IRS denied the taxpayer the normally allowable deductions.[110]

In 1957, many years subsequent to the passage of § 882(c)(2), the IRS issued regulations pertaining to this rule but did not require that a tax return be filed by a set time.[111] Over three decades later, the IRS again issued regulations pertaining to this rule; this time, however, in order for a taxpayer to secure the deductions allowable under the Code, the IRS set forth a general filing deadline of eighteen months from the time of the return's due date.[112] The issue before the U.S. Court of Appeals for the Third Circuit was the validity of this regulation that adopted a defined filing deadline.[113]

In commenting on the appropriate deference standard, the Third Circuit cited to the Supreme Court decision in *United States v. Mead Corp.*,[114] declaring that the "*Chevron* deference is appropriate only in situations where 'Congress would expect the agency to be able to speak with the *force of law*.'"[115] Under I.R.C. § 7805(a), the Third Circuit observed that Congress directly invited the IRS to promulgate regulations; that being the case, the agency was delegated the authority to make law.[116] Having laid this groundwork, the Third Circuit then concluded that "the resulting regulation is entitled to the *Chevron* deference if it survives *Chevron's* two prong inquiry."[117]

Under step one of *Chevron*, the Third Circuit decided that the statute was written ambiguously (i.e., I.R.C. § 882(c)(2)'s filing requirement used the phraseology *in the manner prescribed in subtitle F*, but the statute failed to elaborate on whether this phraseology contained a temporal component).[118] In light of this statutory ambiguity, the Third Circuit proceeded to conduct step two of *Chevron* and sought to determine the reasonableness of the IRS's actions.[119] In conducting its reasonableness analysis, the Third Circuit offered the following observation: generally, "[r]ules represent important policy decisions, and should not be disturbed if 'this choice represents a reasonable accommodation of conflicting policies

110. *Swallows Holding*, 515 F.3d at 172.

111. *See* Treas. Reg. § 1.882-4 (as amended in 1990); 22 Fed. Reg. 8377 (Oct. 24, 1957).

112. *See* Treas. Reg. § 1.882-4(a)(3)(i).

113. *Swallows Holding*, 515 F.3d at 164.

114. 533 U.S. 218, 229 (2001).

115. *Swallows Holding*, 515 F.3d at 168 (quoting *Mead*, 533 U.S. at 229).

116. *Id.*

117. *Id.* at 169–70 (citing McNamee v. Dep't of Treasury, 488 F.3d 100 (2d Cir. 2007); Hosp. Corp. of Am. v. Comm'r, 348 F.3d 136 (6th Cir. 2003); Bankers Life & Cas. Co. v. United States, 142 F.3d 973 (7th Cir. 1998); United States v. Cook, 494 F.2d 573 (5th Cir. 1974)).

118. *Id.* at 170.

119. *See id.* at 170–72.

that were committed to the agency's care by the statute . . . .'"[120] With this observation in mind, the Third Circuit upheld the regulation's validity, ruling that it was eminently reasonable since the regulation helped the agency fulfill its oversight responsibilities.[121] In a closing comment, the Third Circuit added that "*Chevron* recognizes the notion that the IRS is in a superior position to make judgments concerning the administration of the ambiguities in its enabling statute."[122]

On numerous other occasions after *Chevron*, the Supreme Court has reiterated the deference that IRS regulations should command and the latitude with which the IRS can craft such regulations. Indeed, in *Mayo Foundation for Medical Education & Research v. United States*,[123] not only did the Supreme Court reiterate the Treasury Department's ability to amend its own regulations if troubled by a court's resolution of an outcome,[124] it clarified the universality of the *Chevron* decision:

> The principles underlying our decision in *Chevron* apply with full force in the tax context. . . . Filling gaps in the Internal Revenue Code plainly requires the Treasury Department to make interpretive choices for statutory implementation at least as complex as the ones other agencies must make in administering their statutes. . . . We see no reason why our review of tax regulations should not be guided by agency expertise pursuant to *Chevron* to the same extent as our review of other regulations.[125]

The *Mayo Foundation* decision represents a culmination of sorts for the *Chevron* deference standard.[126] In instances of statutory ambiguity, *Mayo Foundation* explicitly invites administrative agencies, including the IRS, to promulgate reasonable regulations that support the agency's position (even if the agency previously embraced a contrary position).[127]

---

120. *Id.* at 171 (quoting United States v. Shimer, 367 U.S. 374, 382–83 (1961)).

121. *Id.* at 172.

122. *Id.*

123. *See generally* Mayo Found. for Med. Educ. & Research v. United States, 131 S. Ct. 704 (2001).

124. *Id.* at 712–13.

125. *Id.* at 713 (citations omitted).

126. Note that the Supreme Court's decision in *Mead* endorsed the use of the *Chevron* deference standard but in the context of interpretive tax regulations. *See, e.g.*, *Swallows Holding*, 515 F.3d at 168 (relying, as indicated in the text, on *Mead* for the proposition that *Chevron* controls in this context).

127. The only time that an administrative agency's latitude to craft regulations would not be afforded deference would be in those instances when the statute in question was held to be unambiguous. *See* Home Concrete & Supply, LLC v. United States, 634 F.3d 249, 257 (4th Cir. 2011) (holding that because the statute in question was determined to be unambiguous, *Chevron* deference was not applicable).

In the next section, we explore ways in which the IRS should accept the Supreme Court's invitation and draft treasury regulations that defeat transfer tax avoidance techniques such as sales to grantor trusts, regardless of whether the taxpayer's position was previously sanctioned by the courts.[128]

2. The Latitude Afforded the IRS in Responding to Transfer Tax Avoidance Techniques

In instances when Congress unambiguously provides taxpayers with methods to reduce their transfer tax burdens, the IRS must permit such methods to go unchallenged.[129] Indeed, it would be unconstitutional for an agency that is part of the executive branch to invalidate or overrule an unambiguous statute enacted by Congress.[130] For example, I.R.C. § 2702 permits taxpayers to form qualified personal residence trusts.[131] In terms of transfer tax savings, even if the IRS dislikes the transfer tax savings that these trusts are able to achieve, the agency is at a loss to challenge the viability of such trusts.[132]

In contrast, consider those situations in which Congress has either not spoken or has spoken ambiguously and taxpayers are exploiting such congressional silence or statutory ambiguities. In these cases, via § 7805(a), Congress has delegated to the IRS the ability to draft regulations that curtail such exploitation.[133] In the paragraphs that follow, we suggest regulations that the IRS could draft that would put an end to sales to grantor trusts and possibly other transfer tax exploitation devices as well.

Let's start with sales to grantor trusts. There are at least two different regulations that the IRS could promulgate that would have a significantly chilling effect on taxpayers using this technique.

---

128. *See, e.g.*, Treas. Reg. § 26.2601-1 (as amended in 2010). After experiencing a court defeat in *Simpson v. United States*, 183 F.3d 812 (8th Cir. 1999), the IRS rewrote this particular regulation to largely mirror its litigation position. As one author of this paper observed many years ago, "[i]n effect, rather than continuing to litigate with taxpayers, the government declared victory by regulation." Gans, *supra* note 94, at 746. In effect, deference cases such as *Nat'l Cable & Telecomms. Ass'n v. Brand X Internet Servs.*, 545 U.S. 967 (2005), enable the Treasury Department to rewrite tax regulations in a way that overrules judicial defeats.

129. *See Chevron*, 467 U.S. at 842-43 ("If the intent of Congress is clear, that is the end of the matter; for the court, as well as the agency, must give effect to the unambiguously expressed intent of Congress.").

130. *Cf. id.*

131. *See* I.R.C. §2702 (2006); *see also* McCouch, *supra* note 52, at 99 (discussing special valuation rules).

132. *See Chevron*, 467 U.S. at 842–43.

133. *See supra* Section III.B.1.

The first regulation would expand the application of I.R.C. § 2036. Consider the fact that if a taxpayer transfers property into a trust and retains an income right to the property transferred, such a retained right causes estate tax inclusion of the transferred property.[134] Insofar as sales to grantor trusts are concerned—where the income generated on the property sold generally is used to satisfy the installment payment obligations—how do tax practitioners currently avoid the application of § 2036? They instruct taxpayers to first contribute "seed money" to the trust, wait, and then use this seed money as a down payment to purchase the property owned by the grantor.[135] At least until now, tax practitioners have expressed confidence that adding seed money to the process eliminates § 2036 concerns.[136] However, the IRS has the liberty to append the following provision to the § 2036 regulations: Taxpayers who, in exchange for a promissory note, sell or exchange property to a trust will be considered to have retained an interest in such transferred property until such note is satisfied. This proposed provision is aligned with the underlying purpose of § 2036, which is to bring back into a taxpayer's gross estate those assets in which taxpayers retain either a direct or indirect interest, or both direct and indirect interests in transferred property.[137] In the case of a sales to a grantor trust, there is compelling evidence that the transferred trust property is the most critical resource that sustains installment note payments, and as such, the grantor has obviously retained an interest therein.[138]

The second regulation would target the "bona fide sale" exception to § 2036.[139] Under this exception, the Code nullifies the application of § 2036 "in case of a *bona fide* sale for an adequate and full consideration in money or money's worth."[140] The Code specifically uses the adjective bona fide to

134. *See* I.R.C. § 2036(a).

135. Mulligan, *supra* note 19; Louis A. Mezzullo, *Freezing Techniques: Installment Sales to Grantor Trusts*, 14 PROB. & PROP. 16, 19 (2000).

136. At least two commentators have argued that seed money is not a necessary predicate to avoid I.R.C. § 2036 application. *See* Elliot Manning & Jerome M. Hesch, *Beyond the Basic Freeze: Further Uses of Deferred Payment Sales*, 34 INST. ON EST. PLAN. ¶ 16 (2000).

137. *See* I.R.C. § 2036.

138. As suggested, there is dictum in an old Supreme Court decision, *Fidelity-Philadelphia Trust Co. v. Smith*, 356 U.S. 274, 277 (1958), to the effect that the use of seed money can be used to negate the application of I.R.C. § 2036. Because the relevant portion of the decision in *Fidelity-Philadelphia* is not a holding, but only dictum, the IRS would be free to take a different approach by regulation. *See* Nat'l Cable & Telecomms. Ass'n v. Brand X Internet Servs., 545 U.S. 967 (2005). Were the agency to adopt new regulations, it should withdraw or qualify Rev. Rul. 77-193, 1977-1 C.B. 273 (offering tacit endorsement of the dictum found in *Fidelity-Philadelphia*).

139. *See* I.R.C. § 2036(a).

140. *Id.*

modify the noun sale.[141] As recent cases have concluded, this language suggests that the exception should not be available unless there is a sufficient non-tax purpose for undertaking the transaction.[142] Based on these cases, the IRS should add the following provision to the bona fide sale exception: When there is a sale between the taxpayer and a party who is unrelated to the taxpayer (as defined in § 267(b)), the bona fide sale exception automatically applies; in all other cases, there is a rebuttable presumption that a sale is not bona fide unless the taxpayer is able to present clear and convincing evidence to the contrary (i.e., there was a legitimate business purpose underlying such sale). In the case of most sales to grantor trusts, this bona fide element will be absent, because the motivation of the taxpayer-grantor for entering into this transaction is clearly grounded in transfer tax savings rather than a legitimate business purpose.

If the IRS promulgated the foregoing recommended Treasury regulations, some commentators might assert that the IRS would be overstepping its bounds. However, consider how the agency has recently proposed regulations that are designed to put an end to the use of private annuities as devices to achieve transfer tax savings.

A private annuity is a transaction where a taxpayer exchanges property with another taxpayer (usually a younger member of the transferor's family) in return for an unsecured promissory note requiring periodic payments until a set price is met or the transferor of the property dies.[143] In a series of prior judicial decisions spanning the course of several decades, courts have held that such exchanges should be treated as open transactions, applying the annuity proceeds first against the tax basis of the exchanged

---

141. *See* Estate of Schutt v. Comm'r, 89 T.C.M. (CCH) 1353, 1364 (2005) ("In probing the presence or absence of a bona fide sale and corollary legitimate and significant nontax purpose, courts have identified various factual circumstances weighing in this analysis. These factors include whether the entity engaged in legitimate business operations, whether property was actually transferred to the entity, whether personal and entity assets were commingled, whether the taxpayer was financially dependent on distributions from the entity, and whether the transferor stood on both sides of the transaction.").

142. *See, e.g.*, Estate of Thompson v. Comm'r, 382 F.3d 367, 383 (3d Cir. 2004); Kimbell v. United States, 371 F.3d 257 (5th Cir. 2004); Estate of Bongard v. Comm'r, 124 T.C. 95, 122–23 (2005); Estate of Hillgren v. Comm'r, 87 T.C.M. (CCH) 1008, 1014 (2004); Estate of Stone v. Comm'r, 86 T.C.M.(CCH) 551, 578–79 (2003); Estate of Strangi v. Comm'r, 85 T.C.M. (CCH) 1331, 1336–37, 1343 (2003); Estate of Harper v. Comm'r, 83 T.C.M. (CCH) 1641, 1648 (2002). If I.R.C. § 2036 exception instead read "in case of a sale for an adequate and full consideration in money or money's worth," it would have strongly implied that nonbusiness reasons could motivate such sale and, in those instances, that such transactions would have qualified under this exception.

143. John K. Pierre, *Using Intra-Family Sales in Estate Freezing: The Prospects in the Year 2000 and Beyond for Private Annuities and Self-Cancelling Installments Notes*, 24 S.U. L. REV. 207, 208 (1997); John G. Brant, *A New Look at the Private Annuity*, 22 COLO. LAW. 733 (1993).

asset and regarding the exchanges as taxable when received only after full basis recovery has been achieved.[144] In 1953, the IRS acceded to this position and issued Revenue Ruling 53-239.[145] In 1954[146] and again in 1963,[147] the IRS made two unsuccessful attempts to have Congress adopt provisions that would have triggered, upon property exchange, an immediate tax upon the entire realized gain. In 1969, the IRS issued Revenue Ruling 69-74,[148] which retracted Revenue Ruling 53-239 and declared gain resulting from the private annuity exchange to be taxable to the transferor in a ratable fashion (i.e., each annuity payment would simultaneously constitute a return of basis and a taxable gain).[149] But in a reversal of its own position, in 2006, the IRS issued proposed Treasury Regulation 1.1001-1(j)(1).[150] This proposed regulation overrules existing case law and the agency's own administrative guidance, treating the value of the annuity as an amount realized under I.R.C. § 1001 and triggering immediate taxation of the entire realized gain.[151] By promulgating this Treasury regulation, the IRS has effectively eliminated any opportunity for taxpayers to postpone taxable gains on dispositions associated with private annuities; as a practical matter, the IRS's actions have severely curbed the use of private annuities as an estate planning device.[152]

If the IRS can deliver the death knell to private annuities, there is no reason it cannot do the same for grantor trusts and other aggressive transfer tax savings devices. Aside from the proposed regulations we recommend, there are a host of other Treasury regulations—too numerous to expand upon here—that the IRS could issue that would put the brakes on many such devices.[153] *Chevron* and *Mayo Foundation* have paved the path for the

---

144. *See, e.g.*, Comm'r v. Kann's Estate, 174 F.2d 357, 358–59 (3d Cir. 1949) (holding that an individual's unsecured promise to pay an annuity to another has no fair market value for the purpose of computing capital gain); Lloyd v. Comm'r, 33 B.T.A. 903, 904–05 (1936) (holding that a promise to make future payments has no fair market value until actual payments are made).

145. 1953-2 C.B. 53.

146. H.R. REP. NO. 83-1337, at A286 (1954), *reprinted in* 1954 U.S.C.C.A.N. 4025, 4111; S. REP. NO. 83-1622, at 116 (1954), *reprinted in* 1954 U.S.C.C.A.N. 4629, 4749.

147. H. COMM. ON WAYS & MEANS, 88TH CONG., PRESIDENT'S 1963 TAX MESSAGE 134 (Comm. Print 1963) (statement of the Secretary of the Treasury).

148. 1969-1 C.B. 43.

149. *See id.*

150. Exchanges of Property for an Annuity, 71 Fed. Reg. 61,441 (proposed Oct. 18, 2006).

151. Alan S. Lederman, *Proposed Regulations on the Tax Treatment of Private Annuities Would Generally Make Them Unattractive*, 106 J. TAX'N 175, 175–76 (2007).

152. *See generally id.* (explaining why private annuities are no longer a practical tax-saving device).

153. For example, via Treasury regulations, the IRS should seek to overturn the outcome in *Petter v. Comm'r*, 98 T.C.M. (CCH) 534 (2009). In *Petter*, the Tax Court approved the use of a so-called value definition clause. *Id.* at 544. This clause provided that

IRS to follow; hesitancy on the agency's part to take action is proving costly to the fisc.

C. The Economic Substance Doctrine and Transfer Taxation

When a taxpayer sells an item to a grantor trust (i.e., the taxpayer's alter ego), the transaction on its face appears driven entirely for tax avoidance purposes.[154] Therefore, this naturally raises the question of how the Code should treat these sales and others that lack economic substance from the perspective of the income tax.

By way of background, the IRS and courts have historically denied taxpayers losses, credits, and other benefits otherwise allowable under a literal reading of the Code and regulations if the transactions that gave rise to these tax benefits lacked economic substance.[155] In tax parlance, this gloss on the Code became known as the economic substance doctrine.[156] Despite the IRS's success in using the economic substance doctrine to

---

in the event the value of a sold asset is determined to exceed the selling price, the excess would pass to charity. *Id.* at 537. By passing the excess to charity and qualifying for a gift tax charitable deduction under I.R.C. § 2522, the taxpayer avoids any gift tax imposition. *Id.* at 538. If such a clause is inserted into a sales document, the IRS has no incentive to raise valuation issues on audit, because no gift tax revenue can be generated. Given the deleterious nature of value-definition clauses in terms of undermining transfer tax enforcement, the IRS should craft Treasury regulations that nullify their effect and render them void ab initio as a matter of public policy based on a codification by regulation of the decision rendered in *Comm'r v. Proctor*, 142 F.2d 824, 827 (4th Cir. 1944) (holding a clause designed to defeat gift tax imposition if a tax audit resulted in increased value of the gift to be void against public policy).

154. *See generally supra* Section II.A (describing tax implications regarding sales to grantor trusts).

155. Many commentators pin the origin of the economic substance doctrine to Gregory v. Helvering, 293 U.S. 465 (1935), which states:

> In these circumstances, the facts speak for themselves and are susceptible of but one interpretation. The whole undertaking, though conducted according to the terms of [the statutory provision], was in fact an elaborate and devious form of conveyance masquerading as a corporate reorganization, and nothing else. The rule which excludes from consideration the motive of tax avoidance is not pertinent to the situation, because the transaction upon its face lies outside the plain intent of the statute. To hold otherwise would be to exalt artifice above reality and to deprive the statutory provision in question of all serious purpose.

Id. at 470.

156. *See* Joseph Bankman, *The Economic Substance Doctrine*, 74 S. CAL. L. REV. 5, 9 (2000); David P. Hariton, *Sorting Out the Tangle of Economic Substance*, 52 TAX LAW 235, 245 (1999); David A. Weisbach, *An Economic Analysis of the Anti-Tax Avoidance Doctrines*, 4 AM. L. & ECON. REV. 88, 89 (2002).

defeat abusive tax-minimization strategies, there have been many questions over the years regarding its application.[157]

By adding section 7701(o) to the Code,[158] the Health Care and Education Reconciliation Act of 2010 clarified the contours of the economic substance doctrine. Under I.R.C. § 7701(o), for a transaction to have economic substance, taxpayers must meet a twofold test.[159] First, the transaction must change (apart from Federal income tax effects) the taxpayer's economic position (the objective prong of the test).[160] Second, the taxpayer must have a substantial purpose (apart from Federal income tax effects) in engaging in the transaction (the subjective prong of the test).[161] If a taxpayer fails to meet both the objective and subjective prongs of the economic substance doctrine, the tax benefits afforded under Subtitle A of the Code with respect to a transaction are not allowable.[162] For example, in an endeavor to secure a noneconomic tax loss, if a taxpayer devises a strategy to artificially increase an asset's tax basis, the IRS would be at liberty to invoke § 7701(o) to disallow the putative loss associated with the asset's artificially-inflated tax basis.[163]

Despite the seeming breadth of § 7701(o), there are defined boundaries to its application. The section's legislative history, for example, reveals that Congress did not seek to negate or reclassify those transactions that had long-standing judicial and administrative acceptance, even if the choices engendered in these transactions were driven by comparative tax advantages.[164] Second, § 7701(o)(5)(B) states that the economic substance

---

157. In particular, several courts embraced the view that "a transaction would be respected for tax purposes if it had *either* economic substance or a nontax business purpose." Martin J. McMahon Jr., *Living with the Codified Economic Substance Doctrine*, 2010 TAX NOTES TODAY 158-2 (Aug. 2010) (emphasis added).

158. Health Care and Education Reconciliation Act of 2010, Pub. L. No. 111-152, § 1409, 124 Stat. 1029, 1067–68 (2010) (codified as amended at I.R.C. § 7701(o) (West Supp. 2010)).

159. *See* I.R.C. § 7701(o)(1).

160. I.R.C. § 7701(o)(1)(A).

161. I.R.C. § 7701(o)(1)(B).

162. I.R.C. § 7701(o)(5)(A).

163. Prior to the passage of I.R.C. § 7701(o), the IRS commonly used the economic substance doctrine to defeat taxpayers' tax-minimization schemes that were devoid of economic substance. *See, e.g.*, ACM P'ship v. Comm'r, 157 F.3d 231, 252 (3d Cir. 1998) (denying, on the basis of the economic substance doctrine, a corporate taxpayer's losses associated with a partnership created for the purpose of generating a capital loss to offset the corporation's capital gain). *See generally* Jade Trading, LLC v. United States, 80 Fed. Cl. 11 (Fed. Cl. 2007), *aff'd in part, rev'd in part, vacated in part*, 598 F.3d 1372 (Fed. Cir. 2010) (holding that the contribution of offsetting option positions to a partnership followed by a later property distribution from the partnership to the partners did not have economic substance and, for tax purposes, could therefore not be respected).

164. *See* STAFF OF JOINT COMM. ON TAXATION, DESCRIPTION OF REVENUE PROVISIONS CONTAINED IN THE PRESIDENT'S FISCAL YEAR 2010 BUDGET PROPOSAL 44 (Comm. Print

doctrine should not apply to an individual taxpayer's personal transactions; accordingly, § 7701(o) only applies to those transactions entered into in connection with a trade or business or an activity engaged in for profit.[165]

The limitations associated with the breadth of § 7701(o) seem to signify that it does not apply to gratuitous transfers such as gifts and bequests, which, by their very nature, lack economic substance.[166] Indeed, the statute itself appears to contemplate that a taxpayer may enter into a transaction for the purpose of securing a transfer-tax advantage without running afoul of the provision. Given the courts' rejection of the antecedent case-law doctrine in the transfer-tax context,[167] and the reality that transfers for estate-planning purposes are often inherently tax-driven,[168] the statute's failure to focus on transfer-tax savings is not unexpected. Nevertheless, as a matter of sound tax policy and to expand its effectiveness, it would make sense to alter § 7701(o)'s focus to also include targeting abusive transfer-tax strategies.

Sales of assets to grantor trusts potentially constitute a class of such transactions when the invocation of an expanded § 7701(o) would be particularly compelling. Why? While the sale to grantor trust has one foot in the transfer tax realm and is specifically designed to minimize transfer

---

2009) (offering four examples of permissible tax planning, such as the choice between capitalizing a business enterprise with debt rather than with equity).

165. I.R.C. § 7701(o)(5)(B).

166. *See* Comm'r v. Duberstein, 363 U.S. 278, 285 (1960) (quoting Comm'r v. Lo Bue, 351 U.S. 243, 246 (1956); Bogardus v. Comm'r, 302 U.S. 34, 41 (1937)) (internal quotations omitted) (noting that a gift stems from "disinterested generosity" rather than from "the incentive of anticipated benefit" of an economic nature); *see also* Estate of Cristofani v. Comm'r, 97 T.C. 74, 84 (1991) (indicating that the fact that a gift was motivated by tax concerns did not prevent the taxpayer from enjoying the annual exclusion); Richard M. Lipton, *'Codification' of the Economic Substance Doctrine—Much Ado About Nothing?*, 112 J. TAX'N 325, 329 (2010) ("This means that estate and gift planning transfers, which invariably lack a business purpose, are not affected by the codification of the economic substance doctrine.").

167. *See, e.g.*, Estate of Strangi v. Comm'r, 293 F.3d 279, 281–82 (5th Cir. 2002) (rejecting an IRS argument based on the case law doctrine of economic substance in the transfer tax context). For cases in which the courts did take into account business purpose in the transfer tax context, however, *see, e.g.*, Holman v. Comm'r, 601 F.3d 763, 770 (8th Cir. 2010) (although not relying upon I.R.C. § 7701(o) (the passage of which postdates the fact pattern of this case), the U.S. Court of Appeals for the Eighth Circuit held that for gift tax valuation purposes, because the supposed business entity in question conducted "no 'business,' active or otherwise," the valuation of its interests had to be determined under I.R.C. § 2703(a) because the bona fide business arrangement exception rule under § 2703(b) did not apply); Fisher v. United States, No. 1:08-cv-0908-LJM-TAB, 2010 WL 3522952, at *3–*4 (S.D. Ind. Sept. 1, 2010).

168. *See* Estate of Cristofani v. Comm'r, 97 T.C. 74, 84 (1991) (recognizing that the gift transaction before the court was tax motivated, but nonetheless upholding the tax treatment sought by the taxpayer).

taxes,[169] it has another foot entirely in the income tax sphere insofar as the operative document (i.e., an installment sales agreement between the taxpayer and the trustee of the grantor trust) denotes that the transaction in question is a bona fide economic arrangement between two disinterested parties.[170] It is with respect to this latter "foot" that an expanded § 7701(o) would offer the IRS an opportunity: it might then be able to use this Code section to attack the nonrecognition of gain sought by the taxpayer, even though a possible reading of other sections of the Code might produce a contrary result.[171]

Recall how the typical sale to grantor trust operates. The taxpayer contributes funds to a grantor trust; using these funds as a down payment, the trustee of the grantor trust purchases an appreciating asset from the taxpayer, paying the balance of the purchase price with an installment note. The trust earns income, and the taxpayer remains liable for the tax upon that income. On the installment note, the trustee of the grantor trust makes principal and interest payments until the note is fully satisfied.[172] An expanded § 7701(o) could be a highly effective weapon in the IRS arsenal in combating these abusive transactions.

When it comes to revising the economic substance doctrine (now codified in § 7701(o)) to defeat taxpayer transactions that intertwine gratuitous transfers with business transactions, there exists historical precedence. For example, consider the long lineage of gift-leaseback jurisprudence.[173] The facts in these cases often followed a common pattern: Taxpayers establish non-grantor trusts with their children as beneficiaries; taxpayers then gift title to real property to these trusts; these trusts, in turn, lease the contributed real property back to the taxpayers.[174] The whole

---

169. *See* John B. Huffaker & Edward Kessell, *How the Disconnect Between the Income and Estate Tax Rules Created Planning for Grantor Trusts*, 100 J. TAX'N 206, 206 (2004).

170. *See id.* at 210.

171. The agency has unfortunately drawn flawed conclusions (epitomized in Rev. Rul. 85-13, 1985-1 C.B. 184) as to what should be the appropriate tax outcome. As previously discussed, *see supra* notes 63–70 and accompanying text, the IRS should revoke Rev. Rul. 85-13 and thereby remove its tacit endorsement that transactions between taxpayers and grantor trusts fail to give rise to any income tax implications. Alternatively, the IRS should issue notice that with respect to those transactions deemed to lack economic substance, the agency will no longer be bound by Rev. Rul. 85-13.

172. *See supra* Section II.A (explaining the foundation and advantages of how grantor trusts operate).

173. *See, e.g.*, Scott W. Brinkman, *Gifts and Leasebacks: Is Judicial Consensus Impossible?*, 49 U. CIN. L. REV. 379 (1980); G. L. Cohen, *Transfers and Leasebacks to Trusts: Tax and Planning Considerations*, 43 VA. L. REV. 31 (1957); Robert J. Peroni, *Untangling the Web of Gift-Leaseback Jurisprudence*, 60 MINN. L. REV. 735 (1984); Rona J. Rosen, *Gift-Leaseback Transactions: An Unpredictable Tax-Savings Tool*, 53 TEMP. L.Q. 569 (1980).

174. In tax parlance, these particular trust vehicles were known as *Clifford* trusts,

purpose of this arrangement was to enable contributing taxpayers (whose income was subject to high marginal tax rates) to secure rent deductions and assign the corresponding rental income to related taxpayers, namely, their children, whose income was subject to low marginal tax rates.[175]

On numerous occasions, the IRS challenged the validity of such gift-leaseback arrangements.[176] In several instances, the IRS's position was upheld: Courts ruled that because income assignment was the force driving these arrangements, the transaction in question failed the business purpose test.[177] In many other instances, these arrangements were held to be bona fide, and taxpayers were allowed deductions for their rental payments.[178] What the sale-leaseback lineage of cases signifies is that the IRS would be within its historical prerogative to use an expanded § 7701(o), as it has previously used the business purpose doctrine to challenge gratuitous transfers cloaked as legitimate business transactions.

With respect to sales to grantor trusts, if the IRS invoked an expanded § 7701(o), the effects would be salutary. Taxpayers are currently at liberty to engage in such sales in ways that enable them to manipulate the tax system to their advantage. Aside from the sale of appreciating assets to a grantor trust, for example, consider the flexibility that taxpayers enjoy when the

---

eponymously named after *Helvering v. Clifford*, 309 U.S. 331 (1940). Typically, these trusts were irrevocable with terms of ten years and two days and with property reverting back to the taxpayer after the designated trust term lapsed.

175. *See* I.R.C. § 1 (2006) (progressive tax rate structure). *But see* I.R.C. § 1(g) (2006) (taxing unearned income of children under age 19 at their parents' highest marginal tax bracket).

176. *See* Rev. Rul. 54-9, 1954-1 C.B. 20, *modified on other grounds*, Rev. Rul. 57-315, 1957-2 C.B. 624, stating thus:

> Accordingly, it is held that the transfer of real property to a trust for a 10-year period for the benefit of grantor's children with his wife as one of two trustees, with the corpus to go to the grantor's wife in the event of his death prior to the expiration of a 10-year period, and with a privilege of leasing back such property from the trustees constitutes a transfer in form rather than substance. Rental payments made to the trust by the grantor will not constitute deductible business expenses. The grantor will remain the owner of the property during the term of the trust for purposes of Federal income and gift taxes, and the rental payments when made will constitute gifts.

Rev. Rul. 54-9, 1954-1 C.B. 20

177. *See, e.g.*, Matthews v. Comm'r, 520 F.2d 323, 324 (5th Cir. 1975) (sale-leaseback lacked economic reality and thus the rent deductions associated with the lease were disallowed); Perry v. United States, 520 F.2d 235, 236 (4th Cir. 1975) (sale-leaseback transaction was ignored because it lacked business purpose).

178. *See, e.g.*, Brown v. Comm'r, 180 F.2d 926, 929 (3d Cir. 1950) (rent payments constituted legitimate business obligations and as such, were deductible); Skemp v. Comm'r, 168 F.2d 598, 600 (7th Cir. 1948).

terms of an irrevocable trust that has grantor trust status no longer suit the taxpayer's needs or desires. In those circumstances, a taxpayer can establish a new irrevocable trust that also has grantor trust status with more favorable terms and have its trustee purchase the prized assets of the irrevocable trust with the unfavorable terms.[179] Despite the lack of economic substance engendered by such transactions, such transactions (and others like them) have remained shielded from taxation to date.[180] The addition of an expanded § 7701(o) to the IRS's arsenal of weapons would presage the possible end of these taxpayer-friendly outcomes.

As indicated, in the sphere of transfer taxation, § 7701(o) currently has limited application. Congress should consider expanding the application of § 7701(o) to make it applicable to abusive transfer tax arrangements. Admittedly, distinguishing abusive from non-abusive transactions will not be easy. Nevertheless, a sale to a grantor trust arrangement should readily fall on the abusive side of the line.

IV. WHAT CONGRESS CAN DO TO MAINTAIN THE INTEGRITY OF THE TRANSFER TAX SYSTEM

Even if the IRS does its job and monitors tax compliance and the courts do their job in adjudicating disputes between the IRS and taxpayers, the transfer tax system will fail to achieve its intended goals of curtailing inherited wealth and raising tax revenue,[181] unless Congress plays a more active role in promoting the integrity of the transfer tax system. First, Congress should eliminate absurdities that make the transfer tax system appear farcical to ordinary taxpayers. Second, when Congress learns of a statutory flaw or oversight, it should act with alacrity to remedy the problem.

A. Eliminating Absurdities from the Transfer Tax System

Many Code provisions are designed to make the tax system more administrable, efficient, and equitable.[182] In theory, these provisions are grounded in logic and common sense; however, in practice, some of these very same tax provisions have spawned elaborate estate planning techniques that have shrouded the transfer tax system with absurdities, thus

179. *See* Rev. Rul. 2007-13, 2007-1 C.B. 684.

180. *See, e.g.*, I.R.S. Priv. Ltr. Rul. 2005-18-061 (May 6, 2005); I.R.S. Priv. Ltr. Rul. 2005-14-001 (Apr. 8, 2005); I.R.S. Priv. Ltr. Rul. 2002-47-006 (Nov. 22, 2002).

181. *See, e.g.*, James R. Repetti, *Democracy, Taxes and Wealth*, 76 N.Y.U. L. REV. 825, 825 (2001).

182. STANLEY S. SURREY, PATHWAYS TO TAX REFORM: THE CONCEPT OF TAX EXPENDITURES 25–27 (1973).

subverting its public stature. The subsections below highlight three such emblematic provisions.

1. The Grantor Trust Rules

Taxpayers who engage in sales transactions with grantor trusts often marvel at the fact that the Code sanctions their use in ways that produce tremendous transfer tax savings. At the core, what must truly astonish taxpayers is that they can sell appreciated property to an irrevocable trust in which they lack any meaningful indicia of control (hence, such trust assets are not part of their gross estate); however, due to the antiquated grantor trust rules, the taxability of such transactions is ignored.[183] The stunning tax savings that such sales produce cast a harsh light on the interrelationship of the income and transfer tax systems,[184] strongly beckoning Congress to take remedial action.

On the one hand, Congress should consider large-scale reform and attack the root of the problem. The Tax Reform Act of 1986 compressed the income tax bracket structure for both estates and trusts[185] and also introduced the "kiddie tax," a system whereby the unearned income of minor taxpayers is essentially taxed at their parents' highest marginal tax rates.[186] Together, bracket compression and the kiddie tax have largely eliminated the need for the grantor trust rules (which, in large part, were designed to curtail taxpayers' ability to assign income to other taxpayers, such as estates, trusts, and minor children whose incomes, at least in the past, were generally subject to lower marginal tax rates).[187] Combine this obsolescence with the fact that the grantor trust rules are now being used as devices to defeat taxpayers' transfer tax obligations, and what becomes evident is that Congress should take decisive action and repeal these rules,[188] retaining them only in those instances when the trust in question is revocable.[189]

On the other hand, if Congress were to lack the courage or the political will to scrap the grantor trust rules in their entirety, it could institute limited reform measures. More specifically, Congress could amend the Code to provide that all sales between taxpayers and grantor trusts constitute taxable

---

183. *See supra* Section II.A.

184. *See* Danforth, *supra* note 7, at 546.

185. *See* Tax Reform Act of 1986, Pub. L. No. 99-514, § 101(a), 100 Stat. 2085, 2096–97.

186. *Id.* § 1411(a) (enacting I.R.C. § 1(g)).

187. *See, e.g.*, Roswell Magill, *What Shall Be Done with the Clifford Case?*, 45 COLUM. L. REV. 111 (1945) (explaining taxpayers' motivations for utilizing trusts as an income tax-savings device).

188. This proposal largely mirrors a proposal recommended by another commentator. *See* Ascher, *supra* note 41, at 888.

189. *Id.* at 930.

events.[190] By imposing a tax friction on such transactions, most taxpayers would no longer use sales to grantor trusts as a device to circumvent their transfer tax obligations.

2. The Annual Exclusion for Present Interest Gifts

In order to avoid having small or token gifts subject to gift tax, Congress instituted the annual gift exclusion found in I.R.C. § 2503(b).[191] Under this Code section, "present interest gifts" are excluded from gift tax, and the donor need not file a gift tax return.[192] To qualify for this exclusion, such gifts cannot exceed a specified dollar threshold, which is adjusted annually for inflation (in 2011, this dollar amount is $13,000).[193]

To illustrate the mechanics of the present interest rule, consider the following two fact patterns. In 2011, if a mother gifts $13,000 to her daughter, no gift tax is due, and the mother need not file a gift tax return. Suppose instead that the mother makes an identical $13,000 gift, but this time she places the cash in trust for her daughter's benefit. Under these circumstances, the present interest gift tax exclusion would not apply (i.e., the gift into a trust constitutes a future interest),[194] and the mother would have to file a gift tax return and either use a portion of her lifetime gift tax exemption (currently, $5 million)[195] or, if the mother's lifetime gift tax exemption were exhausted, pay gift tax.[196]

Taxpayers have not responded idly to the present interest exclusion. Under the terms of most inter vivos trusts that are irrevocable, taxpayers have fashioned a window period of withdrawal (usually thirty days), during which trust beneficiaries can withdraw contributed gifts.[197] As long as notice of this window period is given, this window period of withdrawal

---

190. This adjustment could most likely be done by adding a new section to I.R.C. § 1001. This subsection would direct that transactions between taxpayers and all grantor trusts, except those trusts that are revocable, would be recognized.

191. I.R.C. § 2503(b) (2006); H.R. REP. NO. 72-708, at 29–30 (1932); S. REP. NO. 72-665, at 41 (1932).

192. I.R.C. § 6019 (2006).

193. I.R.C. § 2503(b)(2).

194. *See* Treas. Reg. § 25.2503–3(a) (as amended in 2011) ("'Future interest' is a legal term, and includes reversions, remainders, and other interests or estates, whether vested or contingent, and whether or not supported by a particular interest or estate, which are limited to commence in use, possession, or enjoyment at some future date or time.").

195. *See* I.R.C. § 2505(a)(1) (2006).

196. *See* I.R.C. § 2501(a)(1) (2006).

197. Malcolm A. Moore, Crummey *Trusts*, *in* 26 PHILIP E. HECKERLING INSTITUTE ON ESTATE PLANNING ¶ 203.1 (John T. Grubatz ed., 1992); Bradley E.S. Fogel, *The Emperor Does Not Need Clothes—The Expanding Use of "Naked"* Crummey *Withdrawal Powers to Obtain Federal Gift Tax Annual Exclusions*, 73 TUL. L. REV. 555, 571 (1998); Kent Mason, *An Analysis of* Crummey *and the Annual Exclusion*, 65 MARQ. L. REV. 573, 593 (1982).

appears to transform otherwise future interest trust contributions into present interest gifts qualifying for the annual exclusion. In *Crummey v. Commissioner*,[198] this creative strategy was given a judicial imprimatur of legitimacy; indeed, the eponymously named *Crummey* withdrawal powers are probably the most commonly incorporated feature of virtually every newly minted inter vivos irrevocable trust.[199]

Crummey withdrawal powers operate to readily defeat the underlying purpose of the gift tax annual exclusion—namely, to shelter taxpayers from the administrative inconvenience of having to account for those gifts that are considered de minimis in nature, such as birthday, wedding, and holiday presents. As a result of widespread *Crummey* power usage, vast amounts of wealth escape from the transfer tax base. Furthermore, its usage makes a mockery of the present interest exclusion.[200] Left unchecked, the use of *Crummey* withdrawal powers siphons large sums of dollars from the transfer tax base and casts the transfer tax as a Maginot Line of sorts that can be easily circumvented.

However, a minor legislative change could make *Crummey* withdrawal powers a thing of the past. Simply put, Congress could declare that any and all direct and indirect contributions to irrevocable trusts fail to qualify for the present interest exclusion.[201] Institution of this simple provision would be the death knell for *Crummey* withdrawal powers and simultaneously strengthen the integrity of the transfer tax system.

---

198. Crummey v. Comm'r, 397 F.2d 82 (9th Cir. 1968).

199. *See, e.g.*, Henry B. Greenberg, *Estate and Gift Issues Relating to Irrevocable Trusts*, DITM MA-CLE 1-1 (2009) (explaining how practitioners should incorporate Crummey provisions into irrevocable trusts); L. Henry Gissel, Jr., *Closing Thoughts for This Century on* Crummey *and Other Irrevocable Trusts (Including Insurance Trusts)*, SE35 ALI-ABA 521 (1999) ("*Crummey* clauses are a familiar estate planning device.").

200. Bradley E.S. Fogel, *Back to the Future Interest: The Origin and Questionable Legal Basis of the Use of Crummey Withdrawal Powers to Obtain the Federal Gift Tax Exclusion*, 6 FLA. TAX REV. 189, 193 (2003) ("It cannot be argued that Crummey powers are anything other than a ruse."); John L. Peschel, *Major Recent Tax Developments in Estate Planning*, *in* 33 U.S. CAL. TAX INST. ch. 14, ¶ 1401 (1981) ("[T]he Crummey power, in theory, has a strong legal basis but, in practice, emits an equally strong odor of sham."); Willard H. Pedrick, *Crummey Is Really Crummy!*, 20 ARIZ. ST. L.J. 943, 948 (1988) ("[T]he [*Crummey*] withdrawal right is transparently a flim flam."); Benjamin N. Henszey, Crummey *Power Revisited*, TAXES: THE TAX MAGAZINE, FEB. 1981, AT 76, 77 ("[T]he IRS is aware that the [*Crummey*] power is a sham in most cases."); DEP'T OF THE TREASURY, GENERAL EXPLANATIONS OF THE ADMINISTRATION'S REVENUE PROPOSALS 130 (Feb. 1998), *available at* http://www.treasury.gov/resource-center/tax-policy/Documents/grnbk98.pdf (noting that "the Crummey power is essentially a legal fiction").

201. *Cf.* TASK FORCE ON FED. WEALTH TRANSFER TAXES, REPORT ON REFORM OF FEDERAL WEALTH TRANSFER TAXES 97 (2004), *available at* http://www.abanet.org/tax/pubpolicy/2004/04fwtt.pdf (describing a similar legislative proposal).

3. The Generation-Skipping Transfer Tax Exemption

Current law provides that a generation-skipping transfer (GST) tax applies if certain events transpire (as defined by the Code to be a taxable distribution, a taxable termination, or a direct skip)[202] in which assets pass to "skip persons" (i.e., essentially a transferee who is two or more generations younger than the transferor).[203] In those instances when a GST tax event occurs, GST tax is applicable.[204] The GST tax rate is set to equal the highest marginal estate tax rate.[205]

The legislative purpose behind instituting the GST tax was to defeat attempts by wealthy taxpayers and their families to circumvent the estate tax.[206] As an example, suppose a taxpayer establishes a trust for the lifetime benefit of his child with the remainder to his grandchild. The terms of this lifetime trust provide the child with the following rights and privileges: an income stream, principal distributions in accordance with an ascertainable standard, lifetime and testamentary special powers of appointment, and the annual ability to withdraw the greater of $5,000 or five percent of the trust corpus. Notwithstanding that during the child's lifetime, he could potentially reap rich financial benefits from the trust, the trust's property would not be includable in the child's gross estate upon the child's death for purposes of the federal estate tax.[207] Such transfers and others like them were and continue to be the targets of the GST tax, which imposes a tax in instances (such as the one posited in the above example) when gratuitous wealth transfers are not subject to tax at every generational level.[208]

To curtail the application of the GST tax to only those instances when taxpayers and their families were truly seeking to circumvent their transfer tax obligations in a significant fashion, Congress added a limited exemption to the GST tax.[209] Its application enables taxpayers to transfer a certain dollar amount (currently, $5 million) free of the GST tax.[210] For instance, a taxpayer can make a $5 million gift to a grandchild without incurring a GST tax.[211]

Notwithstanding congressional intentions of providing a limited exception to GST tax application, taxpayers have capitalized upon the GST

---

202. I.R.C. § 2611(a)(1)–(3) (2006).

203. I.R.C. § 2613(a) (2006).

204. I.R.C. § 2641(a) (2006).

205. *Id.*

206. STAFF OF THE JOINT COMM. ON TAXATION, GENERAL EXPLANATION OF THE TAX REFORM ACT OF 1976 564–66 (Comm. Print 1976) (explaining the need to introduce the generation-skipping transfer tax).

207. *See* I.R.C. § 2033.

208. *See supra* notes 200–02 and accompanying text.

209. *See* I.R.C. § 2631(a) (2006).

210. I.R.C. § 2631(c); I.R.C. § 2010(c)(3)(A).

211. *See* I.R.C. § 2631(c); I.R.C. § 2010(c)(3)(A); I.R.C. § 2613(a).

tax exemption in ways that Congress probably never envisioned. Indeed, taxpayers exploiting this exemption can pass wealth, free of GST tax, not only to so-called "skip" people who are two generations below them (such as grandchildren) but also to much more distant generations such as great-grandchildren and great-great-grandchildren.[212] This ability to pass property to very distant generations without the application of a GST tax combined with the eradication by most states of their rules against perpetuities has given rise to an era of dynasty trust formation.[213] Utilizing the GST tax exemption, wealthy families can now establish trusts, funded with millions of dollars, which are essentially insulated from transfer tax for possibly centuries and millenniums to come.[214]

But there is a relatively easy fix to the dynasty trust problem. Congress can limit GST tax exemption allocation to those instances in which the property in question vests with a skip person not more than two generations below the transferor (i.e., the transferor's grandchildren). If property vests or could vest with a skip person more remote (e.g., great-grandchildren), Congress could prohibit taxpayers from making a valid GST exemption allocation.[215] By narrowing the application of the GST tax exemption in this fashion, Congress would protect the transfer tax base and help eradicate the wealth disparities that dynasty trusts generate.[216]

In sum, the three devices summarized above—use of grantor trusts, *Crummey* withdrawal rights, and dynasty trusts—do not exhaust the field of tax absurdities, but they are representative of the systemic problems inherent in the Code. These absurdities generate taxpayer cynicism, which

---

212. See I.R.C. § 2631(c); I.R.C. § 2613(a).

213. *See, e.g.*, Lawrence W. Waggoner, *Message to Congress: Halt the Exemption for Perpetual Trusts*, 109 MICH. L. REV. FIRST IMPRESSIONS 23, 23–24 (2010) (decrying the use of perpetual trusts by wealthy taxpayers); Robert H. Sitkoff & Max M. Schanzenbach, *Jurisdictional Competition for Trust Funds: An Empirical Analysis of Perpetuities and Taxes*, 115 YALE L.J. 356, 410–11 (2005) (finding that roughly $100 billion in trust assets had flowed into trusts governed by state law that allow perpetual or near-perpetual trusts and that impose no state income tax on trust income produced by funds originating from out of state).

214. *See* Waggoner, *supra* note 213, at 23–24.

215. *See* STAFF REPORT OF JOINT COMM. ON TAXATION, OPTIONS TO IMPROVE TAX COMPLIANCE AND REFORM TAX EXPENDITURES 393, *available at* http://www.house.gov/jct/s-2-05.pdf (Several years ago, the Joint Committee on Taxation made this recommendation.).

216. *See, e.g.*, Ray D. Madoff, *America Builds an Aristocracy*, N.Y. TIMES July 10, 2010, *available at* http://www.nytimes.com/2010/07/12/opinion/12madoff.html (bemoaning the surge of dynasty trust use in United States). Another proposal worthy of consideration would be for Congress to institute a federal law adopting the rule against perpetuities. *See* Mitchell Gans, *Federal Transfer Taxation and the Role of State Law: Does the Marital Deduction Strike the Proper Balance?*, 48 EMORY L.J. 871, 879 (1999) ("To remedy [the problem of perpetual trusts], a federal rule limiting the duration of exempt trusts, . . . would be necessary.").

in turn causes taxpayers to believe that the whole transfer tax system is rife with corruption, i.e., nothing more than a charade in which only those "not in the know" are forced to participate.

B. Act with Alacrity to Remedy Legislative Shortcomings and Oversights

One of the most interesting things about the problems confronting the transfer tax system is that they are hiding in plain sight. For example, virtually any practitioner journal or programming agenda of any estate planning continuing legal education series offers readily available planning devices that exploit legislative shortcomings and oversights.[217] The effects of this publicity are twofold. First, the public discourse and exchange of ideas helps to refine these devices, making them less susceptible to IRS challenge and attack. Second, the very act of publicizing these planning devices provides them with an aura of legitimacy, lending traction and credence to their acceptability.

The availability of this knowledge is in sharp contrast to the tax shelter problem that beset the nation during the 1990s. Consider the fact that during the 1990s, tax practitioners were dispensing numerous putative tax-saving strategies that were costing the nation billions of dollars of lost tax revenue annually.[218] Part of the success of these strategies was that this subterfuge was clandestine;[219] indeed, as part of these arrangements, participating taxpayers were often required to sign nondisclosure agreements. The clandestine nature of these arrangements allowed them to flourish, particularly, because many of these strategies eluded detection upon audit. Once these methodologies were brought to light, the IRS and then the courts agreed that these arrangements lacked economic substance,

---

217. *See, e.g.*, Steve R. Akers, *Estate Planning in 2010 and Beyond: Now What?*, SS007 ALI-ABA 1 (2010); S. Stacy Eastland, *Some of the Best Estate Planning Ideas We See Out There*, CS004 ALI-ABA 29 (2010); Lawrence P. Katzenstein, *Some Interest-Sensitive Estate Planning Techniques (with an Emphasis on GRATS and QPRTS)*, SR042 ALI-ABA 151 (2008); Domingo P. Such, III, *Advance Transfer Tax Planning Issues for Estate Planning for the Family Business Owner*, SS008 ALI-ABA 597 (2010).

218. *See* COMM. ON HOMELAND SEC. & GOV'T AFFAIRS, THE ROLE OF PROFESSIONAL FIRMS IN THE UNITED STATES TAX SHELTER INDUSTRY, S. REP. NO. 109-54, at 11 n.21 (2005), *available at* http://www.gpo.gov/fdsys/pkg/CRPT-109srpt54/html/CRPT-109srpt54.htm (reporting that, according to the General Accounting Office, a recent IRS consultant estimated that between 1993 and 1999, the IRS lost on average between $11 billion and $15 billion each year from abusive tax shelters).

219. *See* Chin-Chin Yap, *The Tax Shelter Game*, 59 TAX LAW. 1021, 1022 (2006) ("Likewise, the results of enforcement efforts to stem the tide of abusive tax shelters are speculative at best in the secretive, elastic, and innovative world of the tax shelter industry.").

largely putting an end to their existence.[220] In the aftermath of the financial damage that such techniques caused the government, Congress subsequently instituted various disclosure measures designed to curb the reoccurrence of such tax shelters and to put the IRS on notice regarding the use of these and similar income tax shelter techniques.[221]

But in the transfer tax sphere, Congress has not exercised this same vigilance. This is evidenced by the fact that Congress has a readily available road map—via practitioner journals and continuing education legal lecture series—of the most utilized transfer tax savings devices. There are a host of reasons why Congress has taken little or no action to defeat such devices, but the primary one is lack of political will. Although the transfer tax system remains an easy target for scorn and ridicule if it continues to be littered with loopholes and silly absurdities such as those described in the prior subsection, constituents are not interested in paying more taxes. This dislike of the transfer tax system has translated into complete inaction toward making it any sounder, allowing it to remain dysfunctional. Indeed, over the past decade, there have been numerous proposals championed to eliminate the transfer tax system in its entirety[222] or to strip it of any effectiveness.[223]

At some future point in time, if Congress wants to make the transfer tax system more effective, it can readily do so by closing publicized loopholes. Establishing a transfer tax oversight commission that reports annually to Congress would be a step in the right direction. Assuming that the commission's recommendations are taken seriously, quick congressional action to close down transfer tax planning strategies will have a significant salutary effect: It will drive practitioners to be less vocal about their planning ideas and exploitation devices, and in the absence of these techniques being tested and refined in the public domain, it will have a tremendous chilling effect upon their use.

In the realm of transfer tax reforms that Congress should undertake, the suggestions outlined above are but a smattering of the plethora of changes

---

220. Karen C. Burke & Grayson M.P. McCouch, *COBRA Strikes Back: Anatomy of a Tax Shelter*, 62 TAX LAW. 59, 60 (2008).

221. Taxpayers are required to disclose the details of reportable transactions in which they participate by filing IRS Form 8886, "Reportable Transaction Disclosure Statement," with the IRS Office of Tax Shelter Analysis in Ogden, Utah. Treas. Reg. § 1.6011-4(d) (2010). At least one commentator has praised the effectiveness of such disclosure requirements. *See* Ronald A. Pearlman, *Demystifying Disclosure: First Steps*, 55 TAX L. REV. 289, 323 (2002) (asserting that disclosure measures are a "powerful tax enforcement tool" leading to "enhanced compliance").

222. *See, e.g.*, Death Tax Repeal Permanency Act of 2005, H.R. Res. 202, 109th Cong. (2005) (calling for the elimination of the transfer tax system).

223. *See, e.g.*, Permanent Estate Tax Relief Act of 2006, H.R. Res. 885, 109th Cong. (2006) (raising the exemption to $5 million and setting the top estate tax rate equal to the capital gains tax rate).

that should be legislatively instituted. Notwithstanding this fact, if the above few recommendations were instituted and an oversight commission was put into place, ordinary taxpayers would most likely take their transfer tax obligations more seriously, resulting in greater taxpayer compliance. Correspondingly, the transfer tax system would become much better positioned to accomplish its intended goals of curtailing inherited wealth and raising revenue.[224]

V. CONCLUSION

As evidenced by statistical data, transfer taxes apply to only the wealthiest slice of taxpayers.[225] These same taxpayers have ample resources to secure professional advice and to devise ways to minimize their transfer tax burdens. And for close to a century,[226] they have received a healthy return on their professional advice investment, reaping huge transfer tax savings.[227]

A sale to grantor trust represents one planning device that has gained traction in the estate planning community and will likely gain in popularity as Congress is possibly poised to put the brake on other techniques. As the popularity of this technique gains momentum, it threatens the vibrancy of the transfer tax system by reducing its capacity to raise revenue and to curb accumulations of inherited wealth.[228] Other transfer tax savings devices play this same destructive role, and new device formulations no doubt loom on the horizon.

In its existing arsenal, the IRS has weaponry at its disposal to defeat these transfer tax savings devices. In particular, under *Chevron*, the Supreme Court has accorded the IRS significant latitude to draft Treasury regulations that can eliminate many of these planning devices.[229] The only question is how far the IRS can go in successfully employing this strategy without generating the perception that it is overreaching.

The strength of statutory language, rather than regulations, however, ultimately dictates the soundness of any tax system. That being the case, Congress must be vigilant and nimble in crafting legislation and should curb taxpayers' ability to game the system. Furthermore, once Congress

---

224. *See* Repetti, *supra* note 181.

225. For most years that the estate tax has been in effect, its application has been limited to two percent of the nation's wealthiest individual taxpayers. Darien B. Jacobson, Brian G. Raub & Barry W. Johnson, *The Estate Tax: Ninety Years and Counting*, IRS PUBLICATIONS, June 22, 2007, at 118, *available at* http://www.irs.gov/pub/irs-soi/ninetyestate.pdf; Jane G. Gravelle, *CRS Updates Estate Tax Options Report*, 2010 TAX NOTES TODAY 114–23 tbl.1 (2010).

226. *See* Act of September 8, 1916, Pub. L. No. 64-271, 39 Stat. 756.

227. *See* Schmalbeck, *supra* note 10.

228. *See* Repetti, *supra* note 181.

229. *See Chevron*, 467 U.S. at 864–66.

learns that taxpayers have devised methods to breach the system, it must react with deliberate speed to close such statutory gaps.[230]

A case study is typically representative of a larger phenomenon and, as such, can be an effective tool in analyzing important policy issues. Examining a sale to grantor trust is such a case study. This technique represents a broad spectrum of transfer tax savings strategies, and the use of this device illuminates those reform measures that are necessary to improve and overhaul the transfer tax system. Like any case study, however, its ultimate effectiveness is determined by whether those with political power heed its lessons and actually apply them.

230. Rhetorical question: After the court decision in *Walton v. Commissioner*, 115 T.C. 589 (2000), *acq.* 2003-2 C.B. 964, which sanctioned the use of zeroed-out GRATs, why has Congress not yet taken action to close this gaping loophole?

PROPERTY RIGHTS, PROPERTY WRONGS, AND CHATTEL DISPOSESSION UNDER SELF-STORAGE LEASES

JEFFREY DOUGLAS JONES*

I. INTRODUCTION: SETTING ASIDE THE CONSTITUTIONAL QUESTION 1015
II. THE WILD, WILD LEASE: SELF-STORAGE AGREEMENTS AND DEFAULT PRACTICES 1017
A. *Owner's Liens* 1020
B. *Lien Attachment and Risk of Loss* 1022
C. *From Default to Lien Enforcement* 1027
D. *Lien Enforcement Transferring Title to Landlord* 1029
E. *Lien Enforcement by Public Auction* 1030
III. PROPERTY WRONGS OF SELF-STORAGE LAW 1032
A. *Tort Law Misfires and the Elimination of Bailments* 1033
B. *Bona Fide Purchases and the Price Function in Commercially Reasonable Sales* 1037
C. *A Parallel to Residential Landlord-Tenant Law* 1040
D. *The "Crap" Rule* 1043
IV. CONCLUSION: PERSONAL ACCUMULATION AND THE PROPERTY ETHIC "POSSESS LESS 1045

I. INTRODUCTION: SETTING ASIDE THE CONSTITUTIONAL QUESTION

Self-storage leases are troubling. Under such leases, self-storage facility owners may freely dispose of defaulting tenants' medical and tax records, family ashes, and heirlooms in the same manner as they would fungible items such as chairs or a bookshelf.[1] Facility owners are legally entitled to do so through facility-sponsored auctions, most of which are unrestricted by any duty to conduct commercially reasonable sales.[2] Still worse, these legal self-storage practices have generated a clandestine culture of treasure-

* Jeffrey Douglas Jones, Associate Professor of Law, Lewis & Clark Law School; J.D., The University of Michigan Law School-Ann Arbor; Ph.D. Philosophy, The University of Wisconsin-Madison. The author would like to thank participants at the American Association for Law, Property and Society Conference (Georgetown Law School, March 4–5, 2011), and my colleagues and Lewis & Clark, for invaluable comments earlier drafts of this article.

1. *See infra* note 20 and accompanying text.
2. *See infra* notes 114–16 and accompanying text.

hunting that often leaves tenants—some of whom default due to medical emergencies, bankruptcy, or homelessness—with little opportunity to either regain good standing with the owner or obtain fair market value for their belongings.

Legal suspicion of self-storage leases is not new. In the early 1980's, Professor Paul Brest questioned the constitutionality of state self-storage legislation authorizing the aforementioned practices.[3] Professor Brest argued that state statutes granting self-storage landlords the power of self-help—that is, the power to seize and sell tenant property—constitute state action subject to the Due Process Clause of the United States Constitution.[4] Brest reasoned that the power to take property is an exclusive state power.[5] Without state delegation of this power to self-storage landlords by statute, self-storage landlords would be limited to existing common law remedies for private creditors—mainly, obtaining a money judgment in court.[6] Because state self-storage legislation ceded the governmental power to take property to self-storage landlords and dramatically altered creditor-debtor relations in favor of creditors, Professor Brest argued that such legislation must satisfy constitutional due process obligations.[7]

Professor Brest's position was a reaction to *Flagg Brothers v. Brooks*,[8] a U.S. Supreme Court case that, unfortunately, held that state self-storage legislation in fact *was not* state action subject to the Due Process Clause of the United States Constitution.[9] Although Justice Stevens' dissent in *Brooks* prefigured Professor Brest's due process argument,[10] Justice Rehnquist's opinion for the majority stands undisturbed.

Unconstitutionality is not the sole, or even the typical, characteristic of bad law, however. More often, legal rules are considered bad when they fail to conform with the settled doctrines and policy rationales of the body of law that the legal rules are meant to express. The constitutional questions raised by self-storage statutes concern creditor-debtor relationships; however, self-storage leases also create *landlord-tenant* relationships, and such leases have yet to be measured against the common law of property.

This Article argues that the practices of the self-storage industry with regard to defaulting tenants violate well-established doctrines of U.S. property law. Part II reviews the standard terms of self-storage agreements

---

3. Paul Brest, *State Action and Liberal Theory: A Casenote on Flagg Brothers v. Brooks*, 130 U. PA. L. REV. 1296, 1323 (1982); *see also* Alan R. Madry, *State Action and the Due Process of Self-Help; Flagg Bros. Redux*, 62 U. PITT. L. REV. 1, 3–4 (2000) (arguing that the New York statute authorizing self-help was state action subject to the Due Process Clause, but that such claims brought against private parties have no place in federal courts).

4. *See* Brest, *supra* note 3, at 1316, 1329.

5. *See id.* at 1327, 1329.

6. *Id.* at 1305.

7. *Id.* at 1312.

8. 436 U.S. 149 (1978).

9. *Id.* at 152–53, 166.

10. *Id.* at 169–79.

and the remarkable imbalance of rights and duties between self-storage facility owners and tenants. Part III reveals four fundamental property wrongs occurring in self-storage law. Part IV evaluates possible legislative and judicial remedies, as well as remedies growing out of a property ethic that supports personal accumulation.

II. THE WILD, WILD LEASE: SELF-STORAGE AGREEMENTS AND DEFAULT PRACTICES

As the number of self-storage facilities in the United States has multiplied, sales of the stored possessions of defaulting tenants occur more frequently. The following passage describes a typical sale by a self-storage facility in Oregon:

> The second that Gary Reuter yanks up the green sliding metal door of a self-storage unit, the pack of hunters turns on its dozen flashlights.
>
> There are 11 men and one woman, all around retirement age. Most wear canvas jackets or windbreakers, their heads topped with wool caps or wide-brimmed felt Western hats. They hold their industrial Black & Decker LED-beam flashlights over their heads and lean into the dark locker, like spelunkers peering into a cave. They bend to the left and right, moving around each other for a better view, taking care not to step over the concrete threshold of the doorway.
>
> They stay outside the door because, at 9:45 am on this Tuesday in November, the contents of Extra Space Storage unit F27 in Hillsboro still belong to Tia Holland.
>
> Reuter begins the spiel he will recite at each of the fifteen lockers he will open this morning.
>
> "Strictly cash only," he begins. "You've got to have the money right here and now. You've got to leave personal items: photos, tax records, yearbooks, Bibles. Everything is sold as is, where is, how is. You've got 24 hours to clean it out. You must have your own lock. If you do not have a lock, you can buy one from me for $20."
>
> Having your own lock is important. Because in five minutes, this storage unit will no longer belong to Holland. It will belong to the highest bidder from the hunters, even if that bid is only $1.[11]

In the United States, self-storage facilities are now a primary locus of personal property second only to home residences. Here are some striking facts:

- "The self-storage industry in the United States [had] a collective $20+ billion in annual revenues in 2010" and is comprised of more than 51,000 facilities.[12]

---

11. Aaron Mesh, *Raiders of the Lost Crap*, WILLAMETTE WEEK, Dec. 17, 2008, *available at* http://www.wweek.com/portland/article-9984-raiders_of_the_lost_crap.html.

12. *Fact Sheet*, SELF STORAGE ASS'N, http://www.selfstorage.org/ssa/Content/

- "There is a self storage space inventory of 19.2 sq. ft. per U.S. household[.]"[13]
- "There is 7.0 sq. ft. of self storage space for every man, woman and child in the nation; thus, it is physically possible that every American could stand—all at the same time—under the total canopy of self storage roofing[.]"[14]
- "It took the self storage industry more than 25 years to build its first billion square feet of space; it added the second billion square feet in just 8 years (1998-2005)[.]"[15]
- "During the peak development years (2004-2005) 8,694 new self-storage facilities (approximately 480 million square feet of space were added)."[16]
- "At year-end 1984 there were 6,601 facilities with 289.7 million square feet (26.9 million square meters) of rentable self storage in the U.S. At year end 2010, there are approximately 46,500 'primary' self storage facilities representing 2.24 billion square feet"—an increase of more than 1.95 billion square feet.[17]

Another striking fact is that across the United States, people who leased self-storage units are losing the personal property stored within their units: cars, tools, clothing, family photos and other heirlooms, tax and medical records, bikes, human skulls, televisions, computers, lawn mowers, the cremated remains of family members, furniture, pornography, and dangerous junk.[18] The loss of personal property affects the working

---

NavigationMenu/AboutSSA/FactSheet/default.htm (last visited Oct. 10, 2011).

13. *Id.*

14. *Id.*

15. *Id.*

16. *Id.*

17. *Id.*

18. *See* Mesh, *supra* note 11; Jon Mooallem, *The Self-Storage Self*, N.Y. TIMES MAG., Sept. 6, 2009, at 24; *All Things Considered: Blind Auctions Help Self-Storage Firms Recoup Losses* (National Public Radio radio broadcast May 25, 2009) (transcript available at http://www.npr.org/player/v2/mediaPlayer.html?action=1&t=1&islist=false&id=104521902&m=104521908). Newspaper coverage of self-storage dispossession has been widespread. *See, e.g.*, Becky Bartkowski, *Storage-Unit Auction Yields Human Skull in Box*, ARIZ. REPUBLIC, Aug. 9, 2007, *available at* http://www.azcentral.com/arizonarepublic/local/articles/0809phxskull0809.html?&wired; Dan Bernstein, *Their Lives in Storage*, PRESS-ENTERPRISE (Riverside, Cal.), Mar. 28, 2009, *available at* http://nl.newsbank.com/nl-search/we/Archives?p_product=RS&p_theme=rs&p_action=search&p_maxdocs=200&s_dispstring=%22their%20lives%20in%20storage%22&p_field_advanced-0=&p_text_advanced-0=(%22their%20lives%20in%20storage%22)&xcal_numdocs=20&p_perpage=10&p_sort=YMD_date:D&xcal_useweights=no; Mark Boshnack, *Couple Search for Urn's Owner*, DAILY STAR (Oneota, N.Y.), May 19, 2009, *available at* http://thedailystar.com/local/x112915145/Couple-search-for-urns-owners; Clayton Collins, *You Store It, You Lock It, You Stop Paying, You Forfeit It*, RECORD (Bergen County, N.J.), Nov. 2, 2006, at F05; Kevin DeMarrais, *Buying Blind: Bidding on Items Left in Self-Storage Is Boom or Bust*,

homeless,[19] the routine down-and-out,[20] and even, as in the case of disgraced Illinois Governor Rod Blagojevich, the rich and infamous.[21] Denos Communications, a company that advertises auctions for some of the largest self-storage companies in the United States, estimates that in California alone, $3,000,000 is exchanged at 9,000 self-storage auctions every year, with auctions occurring at 800 of California's nearly 3,000 facilities every month.[22]

These property losses occur because every state except Alaska has passed self-storage lien laws.[23] These laws provide self-storage facility owners with robust lien security interests in any personal property placed in

---

RECORD (Bergen County, N.J.), Aug. 6, 2009, at L07; Kim Fassler, *It's a Great Treasure Hunt*, HONOLULU ADVERTISER, Aug. 14, 2007, at B1; Paul Grimaldi, *Looking for Hidden Treasures*, PROVIDENCE JOURNAL-BULLETIN, June 18, 2009, at A; Jessica Heslam, *Late on Payments, Woman Could Lose Belongings in Storage*, BOSTON HERALD, July 2, 2010, at 5; Jeff Kunerth, *Bidding on Leftovers for a Living*, ORLANDO SENTINEL, Sept. 26, 2008, *available at* http://articles.orlandosentinel.com/2008-09-26/news/selfstore26_1_self-storage-units-self-storage-auctions-fencing-business; Karin Price Mueller, *Losing a Lifetime of Belongings: Storage Facility Sells Items When Bill Isn't Paid*, STAR-LEDGER (Newark), Oct. 27, 2009, at 51; Beth Quimby, *Your Records for Sale to the Highest Bidder? Files Abandoned in Self-Storage Can Be Sold Off Like Office Furniture*, PORTLAND PRESS HERALD, Dec. 26, 2009, at A1; Paul Sisson, *OCEANSIDE: Finding Bargains in a Dusty Box*, NORTH COUNTY TIMES (San Diego), Aug. 14, 2009, *available at* http://www.nctimes.com/news/local/oceanside/article_86285338-7642-5812-8057485a125e96b8.html; Danielle M. Williamson, *Urban Treasure Hunting: Bidders Strive to Separate Gold from Dross at Storage Unit Auctions*, WORCESTER TELEGRAM & GAZETTE, Dec. 29, 2008, at B1.

19. *See* Ric Kahn, *Homeless Strain to Keep a Roof Over Their Stuff*, BOSTON GLOBE, July 15, 2007, *available at* http://www.boston.com/news/local/massachusetts/articles/2007/07/15/homeless_strain_to_keep_a_roof_over_their_stuff/.

20. Delores Flynn, *Bad Economy Fuels Storage Unit Auction Boom*, DETROIT NEWS, June 23, 2008, at A1; Lou Hirsh, *Stored Stuff Getting Left Behind*, PRESS-ENTERPRISE, Dec. 30, 2008, *available at* http://nl.newsbank.com/nl-search/we/Archives?p_action=list&p_topdoc=21; Bryn Mickle, *Somebody Is Losing Their Stuff Every Day: Foreclosure Crisis Boosts Storage Units—and Auctions*, FLINT J., May 17, 2008, at 1; Waveney Ann Moore, *Abandoned Self-Storage Units Another Sign of the Times*, TAMPA BAY TIMES, July 4, 2008, *available at* http://www.tampabay.com/news/humaninterest/article661881.ece; Arlene Satchell, *An Industry in Flux: Housing, Job Crises Forcing More to Abandon Property Left at Public Storage Units*, SUN SENTINEL, Mar. 8, 2009, *available at* http://articles.sun-sentinel.com/2009-03-08/business/0903060257_1_storage-units-self-storage-public-storage.

21. Jo Napolitano, *Blagojevich's Elvis Statue Has Left the Storage Facility*, CHICAGO TRIBUNE, Aug. 19, 2010, *available at* http://articles.chicagotribune.com/2010-08-19/news/ct-met-blagojevich-storage-auction-20100819_1_storage-facility-governor-blagojevich-cell-phone.

22. *See About Self Storage Auctions*, SELF STORAGE CAL., http://www.storageauctions.com/3.htm (last visited Sept. 7, 2011).

23. For brief summaries of state-by-state self-storage lien laws, *see* SELF STORAGE LAWS, http://www.storagelaws.net/ (last visited Sept. 1, 2011) and SELFSTORAGES.COM, http://www.selfstorages.com/ (last visited Sept. 1, 2011).

self-storage units.[24] The national consistency of self-storage lien laws is not a coincidence. The self-storage industry has a powerful national lobby, the Self-Storage Association (the "SSA"):

> The SSA advocates for the self-storage industry at the federal level, at the state level (working with our affiliated state associations), and at the local level when necessary. The Association fights to have state-of-the-art lien laws in place, efficient and streamlined lien notification processes, licensing for offering tenant insurance, and adequate late fees. The SSA fights against the federal government installing self-storage monopolies on military bases, against state and local sales taxes on self-storage rents and for reasonable property taxes. The SSA fights for tax reform that will quickly free up storage space in federal bankruptcy actions, and for reasonable abandoned records management, disposition and other privacy issues.[25]

The result of this lobbying, described in the next section, is a largely uniform set of state self-storage laws that grants facility owners broad rights, such as control over a tenant's stored property in the event of default, but grants relatively few protections for tenants themselves.

A. Owner's Liens

Every state except Alaska has a statute that governs property placed in self-service storage facilities.[26] The "Self-Service Storage Act" and the "Self-Service Facility Storage Act" are common short titles for these statutes.[27] All of the statutes grant owners of self-storage facilities an "owner's lien" upon all personal property placed within the facilities by tenants.[28] The owner's lien provision is boilerplate in forty-five of the forty-

---

24. *Id.*

25. *Legislative and Regulatory Resources*, SELF STORAGE ASS'N, http://www.self storage.org/ssa/Content/NavigationMenu/Resources/LegislativeRegulatory/default.htm (last visited Sept. 1, 2011).

26. *See supra* note 23 and accompanying text.

27. *See, e.g.*, ALA. CODE § 8-15-30 (2002); DEL CODE. ANN. tit. 25, § 4901 (2009); VA. CODE ANN. § 55-416 (2007); W. VA. CODE § 38-14-1 (2005).

28. *See, e.g.*, ALA. CODE § 8-15-33 (2002); ARK. CODE ANN. § 18-16-402(a) (2003); CAL. BUS. & PROF. CODE § 21702 (West Supp. 2011); COLO. REV. STAT. § 38-21.5-102 (2010); CONN. GEN. STAT. § 42-160 (2007); D.C. CODE § 40-403 (2001); DEL. CODE ANN. tit. 25, § 4903 (2009); FLA. STAT. § 83.805 (2004); GA. CODE ANN. § 10-4-212 (2009); HAW. REV. STAT. § 507-62 (2006); IDAHO CODE ANN. §55-2305 (2007); 770 ILL. COMP. STAT. 95/3 (2001); IND. CODE ANN. § 26-3-8-11 (2005); IOWA CODE § 578A.3 (1992); KAN. STAT. ANN. § 58-816 (2005); KY. REV. STAT. ANN. § 359.220 (LexisNexis 2008), LA. REV. STAT. ANN. § 9:4758 (2007); ME. REV. STAT. 10, § 1374 (2009); MD. CODE ANN., COM. LAW § 18-503 (2005); MASS. ANN. LAWS 105a § 3 (1999); MINN. STAT. § 514-972 (2002); MO. REV. STAT. § 415.415(1) (2011); NEV. REV. STAT. ANN. § 108.4753(1) (2009); N.H. REV. STAT ANN. § 451-C:2 (2002); N.J STAT. ANN. § 2A: 44-189 (2000); N.M. STAT. ANN.

nine states with statutes governing self-storage facilities.[29] These provisions all contain language similar to that found in Alabama's Self-Service Storage Act:

> [T]he owner of a self-service storage facility and his heirs, executors, administrators, successors, and assigns shall have a lien upon all personal property located at a self-service storage facility for rent, labor, or other charges, present or future, in relation to the personal property and for expenses necessary for its preservation or expenses reasonably incurred in its sale or other disposition pursuant to this article.[30]

The remaining four states with statutes governing self-storage facilities —Montana, Texas, Wyoming and Nebraska—have self-storage owner's lien provisions that use different language to similar effect.[31] For example, Montana law provides that "[a] person who rents storage space to another may sell at public auction the contents of the storage space if the owner of the contents is more than 30 days in default in paying rental fees on the space."[32] Texas self-storage law contains this simple statement: "A lien under this chapter attaches on the date the tenant places the property at the self-service storage facility."[33] Wyoming's personal property law provides that "[a]ny person is entitled to a lien on any goods, chattels, or animals for his reasonable charges for work or services performed . . . ."[34] Furthermore, the law provides that "[a] person engaging in self-storage operations whereby members of the public rent space from the person to store goods and chattels and retain control over access to the goods and chattels . . . is entitled to a lien under this section."[35] In Nebraska, personal property placed in a self-storage facility is governed by the state's Disposition of Personal Property Landlord-Tenant Act.[36]

---

§ 48-11-5 (2011); N.Y. LIEN LAW § 182(6) (McKinney 2007); N.C. GEN. STAT. § 44A-41 (2009); N.D. CENT. CODE § 35-33-02 (2004); OHIO REV. CODE ANN. § 5322.02 (2004); OKLA. STAT. 42, § 196A (2001); OR. REV. STAT. § 87.687(1) (2009); 73 PA. CODE § 1904 (2008); R.I. GEN. LAWS § 34-42-3 (1995); S.C. CODE ANN. § 39-20-30 (2010); S.D. CODIFIED LAWS § 44-14-2 (2004); TENN. CODE ANN. § 66-31-104 (2004); UTAH CODE ANN. § 38-8-2 (2010); VT. STAT. ANN. CODE 9, § 3904 (2010); VA. CODE ANN. § 55-418 (2007); WASH. REV. CODE § 19.150.020 (2010); W. VA. CODE § 38-14-3 (2005); WIS. STAT. § 704.90(3)(a) (2008).

29. *See* sources cited *supra* note 28.

30. ALA. CODE § 8-15-33 (2002).

31. *See* MONT. CODE ANN. § 70-6-420(1) (2009); TEX. PROP. CODE ANN. § 59.006 (2007); WYO. STAT. ANN. § 29-7-101(a)–(b) (2011); NEB. REV. STAT. § 69-2302 (2009).

32. MONT. CODE ANN. § 70-6-420(1) (2009).

33. TEX. PROP. CODE ANN. § 59.006 (2007).

34. WYO. STAT. ANN. § 29-7-101(a)–(b) (2011).

35. *Id.*

36. NEB. REV. STAT. §§ 69-2302(1)–2307(2) (2009).

B. Lien Attachment and Risk of Loss

Not all self-storage statutes specify when an owner's lien attaches.[37] The statutes that do address lien attachment provide that the lien attaches at various times, including when a storage rental agreement is entered into,[38] the date upon which rent is unpaid and due,[39] the date personal property is brought to the facility,[40] the date the occupant is in default,[41] or the date specified in a preliminary notice of default.[42]

Many states' self-storage statutes expressly place the risk of loss or damage to stored property wholly on the tenant.[43] A few states require landlords to exercise ordinary or reasonable care.[44] However, statutory allocation of risk is unnecessary because standard self-storage rental agreements invariably place all risk of loss or damage upon tenants.[45]

Consider the self-storage behemoth, Public Storage. Public Storage operates "over 2,200 unique and diverse company-owned locations in the United States and Europe, totaling more than 135 million net rentable square feet of real estate."[46] With over $1 billion in annual revenues, Public Storage trades on the New York Stock Exchange and is a member of the

---

37. *See, e.g.*, D.C. CODE § 40-403(a)–(b) (2001); HAW. REV. STAT. § 507-62 (2006); MD. CODE ANN. COM. LAW § 18-503 (2005); MONT. CODE ANN. § 70-6-411 (2009); WYO. STAT. ANN. § 29-7-101 (2011).

38. *See, e.g.*, ALA. CODE § 8-15-33 (2002).

39. *See, e.g.*, ARIZ. REV. STAT. ANN. § 33-1703A (2007).

40. *See, e.g.*, GA. CODE ANN. § 10-4-212 (2009).

41. *See, e.g.*, MINN. STAT. § 514.972 (2002).

42. *See, e.g.*, WASH. REV. CODE § 19.150.060 (2010).

43. The statutes of many states contain the following boilerplate language: "Unless the rental agreement specifically provides otherwise, the exclusive care, custody and control of any and all personal property stored in the leased space shall remain vested in the occupant." *See, e.g.*, ARK. CODE ANN. §18-16-405(b)(1) (2003), D.C. CODE § 40-405 (2001), KAN. STAT. ANN. § 58-818 (2005); VA. CODE ANN. § 55-420 (2007), WIS. STAT. § 704.90(4) (2008). Some statutes also provide that "the occupant shall bear all risks of loss or damage to such property." *See, e.g.*, ALA. CODE § 8-15-32 (2002).

44. *See* NEB. REV. STAT. § 69-2306 (2009) ("The landlord shall exercise reasonable care in storing the property but shall not be liable to the tenant or any other owner for any loss unless such loss is caused by the landlord's intentional or negligent act."); OKLA. STAT. tit. 42, 194A (2001) ("The duty of care an owner must exercise with respect to personal property located in a self-service storage facility is ordinary care only."); W. VA. CODE § 38-14-7(a) (2005) ("The owner shall use reasonable care in maintaining the self-service storage facility for the purposes of storage of personal property and may not offer to sell insurance to the occupant to cover the owner's risk or lack of care.").

45. Public Storage, *Lease/Rental Agreement* 3 (Aug. 18, 2010) (on file with the Tennessee Law Review).

46. *About Us*, PUBLIC STORAGE, http://www.publicstorage.com/storage-company-info.aspx.

S&P 500 and the Forbes Global 2000.[47] Public Storage describes itself as "among the largest landlords in the world."[48]

To lease a self-storage unit with Public Storage, "Occupants"—Public Storage's term for lessees—must sign an agreement stating that "[t]he total value of all personal property stored by Occupant is agreed to be less than Five Thousand Dollars ($5,000)."[49] An Occupant also "understands that the Premises are not suitable for the storage of heirlooms or other precious, irreplaceable or invaluable personal property, such as rare books, records, or art, objects for which no immediate resale market exists and objects of special or emotional value to the Occupant."[50]

Regarding risk of loss or damage to stored property, Public Storage's standard self-storage lease agreement contains the following exculpatory clauses:

> 7. Insurance. ALL PERSONAL PROPERTY IS STORED BY OCCUPANT AT OCCUPANT'S SOLE RISK. INSURANCE IS OCCUPANT'S SOLE RESPONSIBILITY. OCCUPANT UNDERSTANDS THAT OWNER WILL NOT INSURE OCCUPANT'S PERSONAL PROPERTY. OCCUPANT IS OBLIGATED UNDER THE TERMS OF THIS RENTAL AGREEMENT TO INSURE HIS/HER OWN GOODS. To the extent Occupant's insurance lapses or Occupant does not obtain insurance coverage for the full value of Occupant's personal property stored in or on the Premises, Occupant agrees Occupant will personally assume all risk of loss. Owner and Owner's agents, affiliates, authorized representatives and employees ("Owner's Agents") will not be responsible for, and Occupant hereby releases Owner and Owner's Agents from any responsibility for, any loss, liability, claim, expense or damage to personal property or injury to persons ("Loss") that could have been insured against (including, without limitation, any Loss arising from the active or passive acts, omission or negligence of Owner or Owner's agents) ("the Released Claims"). Occupant waives any rights to recover against Owner or Owner's Agents for the Released Claims. Occupant expressly agrees that the carrier of any insurance obtained by Occupant shall not be subrogated to any claim of Occupant against Owner or Owner's Agents. Occupant understands that if Occupant elects to obtain the insurance available at the property, the additional amount for such insurance coverage must be included with the monthly payments as noted above. Furthermore, all payments received will be applied as noted above. The provisions of this paragraph will not limit the rights of Owner and Owner's Agents under paragraph 8 Limitation of Owner's Liability; Occupant's Liability. By CLICKING OR PLACING INITIALS HERE ___, Occupant acknowledges that he understands the provisions of this

47. *Id.*
48. *Id.*
49. *See supra* note 45.
50. *Id.*

> paragraph and agrees to these provisions and that insurance is Occupant's sole responsibility.
> 8. Limitation of Owner's Liability; Occupant's Liability. Owner and Owner's agents, affiliates, authorized representatives and employees (collectively called "Owner's Agents") will have no responsibility to Occupant or any other person for any liability, expense, damage to their personal property or injury to them arising out of Owner's active or passive acts, omissions, negligence or conversion unless Owner intentionally and/or in bad faith causes the liability, expense, damage or injury. Occupant agrees that Owner and Owner's Agents' total responsibility for any liability, expense, personal property damage and personal injury will not exceed Five Thousand Dollars ($5,000). Occupant shall defend Owner and Owner's Agents against and pay for any damage to property, injury to persons, or any other liability or expense incurred because of anything Occupant does or fails to do in the Enclosed Space, the Parking Space or surrounding areas, unless Owner intentionally and in bad faith causes such damage, injury or other liability or expense. By CLICKING OR PLACING INITIALS HERE ___, Occupant acknowledges that he has read, understands and agrees to the provisions of this paragraph.[51]

Courts routinely enforce provisions such as the one in Public Storage's standard lease agreement. For example, in *Kane v. U-Haul International, Inc.*, the self-storage landlord failed to notify several tenants of a leak in the facility's roof, which caused water damage to the tenants' property.[52] The tenants sued and sought damages in excess of the storage value limits provided in the contract.[53] The Third Circuit affirmed the lower court's ruling that the landlord's failure to notify the tenants of the leak did not rise to the level of wanton and willful misconduct, the only standard of misconduct which could not be exculpated by the contract.[54]

In *Lathers v. U-Haul Co.*, the tenant stored property in a rented self-storage unit and subsequently became delinquent on his payments.[55] The landlord exercised lien rights on the unit and replaced the tenant's lock with one of its own.[56] The tenant paid the back rent and recovered the unit with all stored items intact.[57] When the tenant returned to his unit several weeks later, there was a tag on his lock stating that the lock was improperly locked.[58] When the tenant opened the unit, he discovered that much of his stored property had been stolen.[59] The tenant sued the landlord for

---

51. *Id.*
52. 218 Fed. App'x 163, 165 (3d Cir. 2007).
53. *Id.*
54. *Id.* at 167.
55. 875 So. 2d 839, 839 (La. Ct. App. 2003).
56. *Id.*
57. *Id.* at 839–40.
58. *Id.* at 840.
59. *Id.*

negligence.[60] Using evidence that the landlord was aware of thefts at the facility both before and after the theft of the plaintiff's belongings, the tenant argued that the landlord was aware of a security problem at the facility when the tenant entered into the contract.[61] The Louisiana Court of Appeal affirmed the lower court's ruling that the tenant bore all risk.[62]

In *Cochran v. Safeguard Self-Storage, Inc.*, another Louisiana Court of Appeal case, tenants leased a unit from the defendant self-storage facility under a lease that provided for the non-liability of the landlord for property stored unless "due to the willful acts of gross negligence of [landlord], his agents, servants, or employees."[63] The lease also excluded all warranties by the defendant-landlord and specified that insurance was the tenants' sole obligation.[64] The tenants' stored property was later destroyed by a fire caused by faulty electrical wiring in a junction box, which was located outside the tenants' leased units but on the landlord's premises.[65] The tenants sued the landlord for negligence.[66] Prior to the fire, the landlord observed flickering lights in one of his storage buildings and called an electrician to detect the source of the problem.[67] The electrician identified and repaired two shorts in an electrical outlet.[68] The electrician did not identify any other electrical problems or indicate to the landlord that there was a problem with the building's wiring.[69] The tenants offered the electrician's visit as evidence that the landlord knew or should have known of a defect in the building's wiring.[70] The Louisiana Court of Appeal, however, affirmed the trial court's grant of summary judgment to the landlord.[71]

In *Whipper v. McLendon Movers, Inc.*, a tenant's stored property was damaged when the water pipes servicing the landlord's automatic sprinkler system burst due to cold weather.[72] The tenant sued the landlord.[73] The Court of Appeals of Georgia affirmed the lower court's grant of summary judgment to the landlord, citing the following provision contained in the lease agreement:

---

60. *Id.*
61. *Id.* at 842.
62. *Id.*
63. 845 So. 2d 1128, 1129 (La. Ct. App. 2003) (citation omitted).
64. *Id.* at 1129.
65. *Id.*
66. *Id.* at 1130.
67. *Id.* at 1129.
68. *Id.*
69. *Id.*
70. *Id.* at 1132–33.
71. *Id.* at 1133.
72. 372 S.E.2d 820, 820 (Ga. Ct. App. 1988).
73. *Id.*

> All personal property brought onto the premises by lessee . . . shall be at the risk of lessee, and lessor shall not be liable for any loss or damages for any reason whatsoever to said property. It shall be the responsibility of lessee to adequately insure any property brought onto the premises, and lessor shall have no duty whatsoever to carry any insurance on property brought onto the premises by lessee.[74]

Self-storage leases also exculpate landlords from liability for damage or loss of tenant property caused by the conduct or misconduct of third parties. In *Arruda v. Donham & Dover Investment Properties, Inc.*, the plaintiff rented a self-storage unit.[75] Another tenant, who had leased a unit nearby, caused a fire within his own unit.[76] The fire spread to the plaintiff's unit and destroyed the plaintiff's property—two automobiles and other personal property.[77] The plaintiff-tenant sued the defendant-landlord in negligence.[78] The landlord argued that as a matter of law it could not be held liable even if the plaintiff's allegations of negligence were found to be true.[79]

To support its argument, the landlord relied on terms in the lease agreement providing that the landlord could not "be held responsible for damage to the plaintiff's property."[80] Other key provisions in the agreement read as follows:

> "Landlord (i.e. Dover) shall not be liable to any tenant (i.e. plaintiff) or any other party for any negligent act or omission of landlord."
>
> "All property stored within the unit by tenant shall be at tenant's sole risk and expense."
>
> "Landlord shall not be liable to tenant for any loss or damage that may be occasioned by or through the act or omission to act of other tenants on the premises or of any other person."
>
> "I understand the provision that states the lessor is not responsible for loss or damage to property in my storage space[.]"[81]

Ultimately concluding that the terms of the lease agreement were "clear and to the point," the court ruled that the landlord could not be held liable for the damage to the plaintiff's property.[82]

---

74. *Id.*
75. No. CV 930520972S, 1994 WL 386092, at *1 (Conn. Super. Ct. July 11, 1994).
76. *Id.* at *1.
77. *Id.*
78. *Id.*
79. *Id.*
80. *Id.* (quoting the parties' lease agreement).
81. *Id.* (quoting relevant provisions of the parties' lease agreement).
82. *Id.* at *1, *4.

Despite these rulings, other courts have found self-storage landlords liable for damage or loss of tenant property. In such cases, liability is usually established by demonstrating that the landlord violated an express term of the particular state's self-storage facility act, such as by failing to give a defaulting tenant proper notice prior to selling that tenant's property at auction.[83] Tenants rarely succeed on the tort exceptions contained in the exculpatory clauses of self-storage leases. When tenants do succeed on such claims, they confront the storage value limitations provided for in the contracts.[84]

C. From Default to Lien Enforcement

Upon default, a self-storage facility owner may deny the defaulting tenant access to his or her personal property stored in the leased unit.[85] Some states permit owners to deny access immediately without any notice to the defaulting tenant while others allow owners to deny access in as few as five to ten days after the date of default.[86] Still other states specify a longer period of default before a facility owner may deny unit access, such as fourteen to thirty days, and then access can be denied only after a notice of default has been sent to the defaulting tenant's last known address.[87]

---

83. *See, e.g.*, Cook v. Pub. Storage, Inc., 761 N.W.2d 645, 672 (Wis. Ct. App. 2008) (affirming a jury verdict that landlord violated self-storage statute by failing to meet the notice requirement when mailed notices were returned as undeliverable and by failing to conduct a commercially reasonable sale); Castetter v. Mr. "B" Storage, 699 A.2d 1268, 1271 (Pa. Super. Ct. 1997) (noting that the landlord violated self-storage statute by failing to give proper notice prior to lien sale and by unauthorized entry into tenant's unit).

84. *See, e.g.*, Sec. Self-Storage v. Pauling, No. 11-09-00103-CV, 2010 WL 3170670, at *3 (Tex. App. Aug. 12, 2010) (affirming lower court opinion holding landlord liable for tenant property disposed of upon landlord's mistaken assumption that tenant had abandoned the property); Dubey v. Pub. Storage, Inc., 918 N.E.2d 265 (Ill. App. Ct. 2009) (invalidating storage value limitation contained in lease).

85. *See, e.g.*, ALA. CODE § 8-15-34(4) (2002); ARK. CODE ANN. § 18-16-405(a) (2003); CAL. BUS. & PROF. CODE § 21705(a)(1) (West 2008); COLO. REV. STAT. § 38-21.5-103(c)(III) (2010); DEL. CODE ANN. tit. 25, § 4904(a)(3)(c) (2009).

86. *See, e.g.*, ARK. CODE ANN. § 18-16-401(1) (2003) ("'Default' means the failure to perform on time any obligation or duty set forth in the rental agreement[.]"); *Id.* § 18-16-405(a) (stating, without restriction or qualification, that "[i]f an occupant is in default, the operator may deny the occupant access to the leased space").

87. *See, e.g.*, ALA. CODE § 8-15-34 (2002) ("(1) No enforcement action shall be taken by the owner until the occupant has been in default continuously for a period of 30 days. . . . (4) The owner shall have the right to deny the occupant access to the leased space and the owner may enter and/or remove the personal property from the leased space to other suitable storage space pending its sale or other disposition."); CAL. BUS. & PROF. CODE § 21703 (West 2008) ("If any part of the rent or other charges due from an occupant remain unpaid for 14 consecutive days, an owner may terminate the right of the occupant to the use of the storage space at a self-service storage facility by sending a notice to the occupant's last known address . . . .").

After a landlord is authorized to deny unit access due to tenant default, only a few states require facility owners to grant tenants access to essential items such as personal papers, health aids, and clothing under a specified dollar value.[88] The majority of statutes grant no such privilege to tenants.[89] Furthermore, a defaulting tenant in most states remains exclusively responsible for all damage and loss of personal property, even after the landlord has removed the tenant's lock and replaced it with its own.[90]

The majority of states prohibit a self-storage landlord from commencing a lien enforcement action until the tenant is in noticed default for a minimum number of days.[91] The minimum number of days varies by state from as many as ninety days in Montana and New Mexico,[92] to as few as ten days in Louisiana, New York, and North Dakota.[93] Several states have set the minimum number of days in the fourteen to thirty day range[94] while others states have chosen the forty-five to sixty day range.[95] Wyoming law, however, appears to permit lien enforcement immediately following notification of default to all persons known to claim an interest in the property.[96]

Texas is the only state which requires a landlord to obtain a court order prior to lien enforcement.[97] California, Nevada, and North Carolina have

---

88. *See, e.g.*, MINN. STAT. § 514.972 (2002) ("The occupant may remove from the self-service storage facility personal papers, health aids, personal clothing of the occupant and the occupant's dependents, and personal property that is necessary for the livelihood of the occupant, that has a market value of less than $50 per item . . . ."); WASH. REV. CODE § 19.150.080(1) (2010) (exempting personal papers and photographs from sale).

89. *See, e.g.*, COLO. REV. STAT. 38-21.5-103 (2010); DEL. CODE ANN. tit. 25, § 4904(e) (2009); FLA. STAT. § 83.806(d) (2004); W. VA. CODE § 38-14-5 (2005).

90. *See, e.g.*, OKLA. STAT. tit. 42, § 195A (2001) ("The owner of a self-service storage facility shall not be liable for damages sustained by an occupant, if any, alleged to result from action taken by the owner to prevent access to the self-service storage facility after the occupant has committed an act of default pursuant to the rental agreement.").

91. *See infra* notes 92–96 and accompanying text.

92. MONT. CODE ANN. § 70-6-411(1) (2009); N.M. STAT. ANN. § 48-11-7(A)(3) (2011).

93. LA. REV. STAT. ANN. § 9:4759(4)(e), (5) (2007); N.Y. LIEN LAW § 182(7) (McKinney 2007); N.D. CENT. CODE § 35-33-05(1)(c)–(d) (2004).

94. *See, e.g.*, ALA. CODE § 8-15-34(1) (2002) (30 days); CAL. BUS. & PROF. CODE § 21705(a) (West Supp. 2011) (14 days); COLO. REV. STAT. § 38-21.5-103(1)(a) (2010) (30 days); DEL. CODE. ANN. tit. 25, § 4904(3)(d) (2009) (30 days); FLA. STAT. § 83.806(2)(c) (2004) (14 days); GA. CODE ANN. § 10-4-213 (2009) (30 days); HAW. REV. STAT. § 507-65(E) (2006) (30 days); 770 ILL. COMP. STAT. 95/4(c)(4) (2001) (14 days); N.C. GEN. STAT. § 44A-43(a) (2009) (15 days); OKLA. STAT. tit. 42, § 197(D)(4) (15 days).

95. *See, e.g.*, CONN. GEN. STAT. § 42-162(5) (2007) (60 days); IDAHO CODE ANN. § 55-2306(1) (2007) (60 days); KAN. STAT. ANN. § 58-817(a) (2005) (45 days); KY. REV. STAT. ANN. § 359.230(1)(a) (LexisNexis 2008) (45 days); W. VA. CODE § 38-14-5(a)(1) (2005) (60 days).

96. *See* WYO. STAT. ANN. § 29-7-105(b) (2011).

97. TEX. PROP. CODE § 59.041(a) (2007).

somewhat similar processes that require the landlord to file a verified complaint to enforce the lien if a defaulting tenant files a declaration in opposition to the lien sale.[98] South Carolina offers a "predistress hearing," but this hearing is available only for lease agreements that do not conform to the model provisions provided for in the statute.[99] South Carolina law explains that "the purpose of the predistress hearing is to protect the occupant's use and possession of property from arbitrary encroachment and to prevent unfair or mistaken deprivation of property."[100] As long as they comply with applicable statutory notice requirements, landlords using lease agreements that conform to South Carolina's model provisions may begin lien enforcement without judicial intervention after fifty days.[101]

D. Lien Enforcement Transferring Title to Landlord

Lien enforcement does not always occur in the form of a sale of personal property.[102] A handful of states permit landlords to retain or destroy the personal property of a defaulted tenant following expiration of the notice period, provided the property has a fair market value below a threshold amount of $1,000, $500, $300, or $100.[103]

Oklahoma's self-storage statutes are interesting in that they expressly give the landowner discretion to determine the fair market value of the defaulting tenant's personal property:

> If the occupant abandons or surrenders possession of the self-service storage facility and leaves household goods, furnishings, fixtures, or any other personal property in the self-service storage facility, the owner may take possession of the property, and if, in the judgment of the owner, the property has no ascertainable or apparent value, the owner may dispose of the property without any duty of accounting or any liability to any party.[104]

---

98. CAL. BUS. & PROF. CODE § 21710 (West Supp. 2011); NEV. REV. STAT. § 108.4765 (2009); N.C. GEN. STAT. 44A-43(b)(2)d (2009).

99. S.C. CODE ANN. § 39-20-47(A)–(B) (Supp. 2010).

100. *Id.* § 39-20-47(B).

101. *Id.* § 39-20-45.

102. *See infra* note 103 and accompanying text.

103. *See, e.g.*, NEB. REV. STAT. § 69-2304 (Supp. 2010) ("A notice given pursuant to section 69-2303 shall contain one of the following statements, as appropriate: . . . (2) 'Because this property is believed to be worth less than one thousand dollars, it may be kept, sold, or destroyed without further notice if you fail to reclaim it within the time indicated in this notice.'"); N.H. REV. STAT. ANN. § 451-C:7(I) (2002) (threshold of below $500 for owner disposal of defaulted tenant property without further notice or auction); OKLA. STAT. tit. 42, § 197.1(A) (2001) (threshold of "no ascertainable or apparent value"); OR. REV. STAT. § 87.691(1) (2009) (threshold of $100 or less); WASH. REV. CODE § 19.150.080(b) (2010) (threshold of below $300).

104. OKLA. STAT. tit. 42, § 197.1(A) (2001).

In contrast to Oklahoma's trust in a landlord's subjective valuation of a tenant's personal property, West Virginia requires a more objective valuation of a tenant's personal property.[105] West Virginia permits landlord destruction of tenant property only if the landlord "can demonstrate by photographs or other images and affidavit of a knowledgeable and credible person that the personal property lacks a value sufficient to cover the reasonable expense of a public auction plus the amount of the self-service storage lien[.]"[106]

E. Lien Enforcement by Public Auction

After expiration of the notice period, a landlord who intends to enforce a lien through a public auction is required to publish an advertisement (typically once a week for two consecutive weeks) of the public sale in a newspaper of general circulation in the city or county where the self-storage facility is located.[107] The uniform content of such ads includes the name of the person on whose account the goods are stored, the space or lot number of the occupant, the time, place and manner of the sale, and the location of the storage facility.[108] Some states require a brief description of the goods to be sold.[109] Most states provide that the sale may not take place sooner than fifteen days after the first publication.[110]

A minority of states requires that a landlord who sells a defaulted tenants' personal property through a public auction conduct such an auction in a "commercially reasonable manner" without defining that term.[111] On the other hand, an overwhelming majority of states presume that

---

105. *See infra* note 106 and accompanying text.

106. W. VA. CODE § 38-14-5(a)(1)(B) (2005).

107. *See, e.g.*, ALA. CODE 8-15-34(7) (2002); CAL. BUS. & PROF. CODE § 21707 (West Supp. 2011); COLO. REV. STAT. § 38-21.5-103(e)(I) (2010); DEL. CODE ANN. tit. 25, § 4904(c) (2009); GA. CODE ANN. § 10-4-213 (2009).

108. *See, e.g.*, ALA. CODE 8-15-34(7)(a)–(b) (2002); ARK. CODE ANN. § 18-16-407(a)(1)(D)–(F) (Supp. 2011); CAL. BUS. & PROF. CODE § 21707 (West Supp. 2011); GA. CODE ANN. § 10-4-213 (2009); NEV. REV. STAT. §108.477(2) (2009).

109. *Compare* ALA. CODE 8-15-34(7)(a) (2002) (requiring a brief description of the goods sold), *and* COLO. REV. STAT. §38-21.5-103(e)(I)(B) (2010) (requiring a "brief and general description of the personal property reasonably adequate to permit its identification . . ."), *with* ARK. CODE ANN. § 18-16-407(a)(1) (Supp. 2011) (no description requirement), *and* WYO. STAT. ANN. § 29-7-105(b) (2011) (notification with description not required).

110. Alabama's code is typical. *See* ALA. CODE 8-15-34(7) (2002); *see also* sources cited *supra* note 107. *But see* ARK. CODE ANN. § 18-16-407(2) (Supp. 2011) (requiring only one advertisement with no content requirements and permitting sale seven days thereafter).

111. *See* ALA. CODE § 8-15-34(13) (2002); CAL. BUS. & PROF. CODE § 21707(b) (West Supp. 2011); HAW. REV. STAT. § 507-66(b) (2006); KAN. STAT. ANN. § 58-817(a)(1) (2005); MINN. STAT. § 336.7-210(a) (Supp. 2011); NEV. REV. STAT. ANN. § 108.477(4) (2009); N.Y. LIEN LAW § 182(7) (McKinney 2007); WASH. REV. CODE § 19.150.080(1) (2010); WIS. STAT. § 704.90(6)7c (Supp. 2010).

compliance with the sale provisions in their self-storage statute constitutes a commercially reasonable sale.[112] Thus, these states' self-storage statutes often include a provision similar to the one in Ohio's self-storage statute that limits a landlord's liability based upon statutory compliance:

> If the owner complies with the requirements for sale under this section, the owner's liability to persons who have an interest in the personal property sold is limited to the balance of the proceeds of the sale after the owner has satisfied his lien.
>
> . . . The owner is liable for damages caused by the failure to comply with the requirements for sale under this section and is liable for conversion for willful violation of the requirements for sale under this section.[113]

Limitation of landlord liability based on statutory compliance comports with my earlier observations that tenants rarely succeed on the tort claim exceptions to self-storage leases,[114] and that landlord liability for damage or loss to tenant property usually turns on demonstration that the landlord violated the express terms of the particular state's self-storage facility act.[115]

Three final observations complete this overview of the enforcement of self-storage laws against defaulting tenants. First, states vary on whether landlords and their agents are permitted to participate in the self-storage auctions that they control.[116] In Washington, for example, "[n]o employee or owner, or family member of an employee or owner, may acquire, directly or indirectly, the property sold pursuant to [the Act] . . . or personal papers and personal photographs disposed of under [the Act]."[117] Conversely, in Ohio, "[a]n owner may buy at any public sale held pursuant to [the Act]."[118]

Second, there is the matter of proceeds. All state statutes require a landlord who has conducted an auction to use the proceeds to satisfy tenant deficiency and reasonable sale expenses, and, in cases of windfall, to hold the balance for the tenant to claim within a specified time.[119] The time a

---

112. *See infra* note 113 and accompanying text.

113. OHIO REV. CODE ANN. § 5322.03(M)(1)–(2) (2004); *see, e.g.*, ALA. CODE §§ 8-15-34(13)–(14) (2002).

114. *See supra* notes 52–82 and accompanying text.

115. *See supra* note 83 and accompanying text.

116. *See infra* notes 117–18. Note that there is a difference between participating in an auction and having the right to control it. Landlords in Alabama, for example, *control* various aspects of the auction, including "[t]he time, place and manner of the sale[,]" whether a tenant's personal property will "be sold singly, in lots or as a whole[,]" whether bids will be sealed or open, and obviously, the sale prices set for particular things or sets of things. ALA. CODE §§ 8-15-34(7)b, (10) (2002).

117. WASH. REV. CODE § 19.150.080(4) (2010).

118. OHIO REV. CODE ANN. § 5322.03(K) (2004).

119. *See, e.g.*, IOWA CODE § 578A.4(8) (2008); MISS. CODE ANN. § 85-7-127(4) (1999); TENN. CODE ANN. § 66-31-105(2)(K) (2004); VA. CODE ANN. § 55-419(E) (2007).

tenant has to collect the balance of an auction of his or her personal property varies greatly by state, from only thirty days in Virginia[120] to three years in Alabama.[121] After the expiration of the period within which a tenant may claim the windfall from a sale of personal property, the proceeds may become the property of the landlord by operation of the self-storage statute,[122] escheat to the county or state,[123] or be disposed of pursuant to the state's unclaimed property statute, which often provides a period after which property is deemed abandoned and becomes the property of the landlord.[124] Arizona and Pennsylvania are unique in requiring such proceeds to go to the Arizona public schools and the Pennsylvania Department of Revenue, respectively.[125]

Finally, some state statutes provide for what should occur when a landlord attempts to auction the personal property of a defaulted tenant, but some or all of the items do not sell. The statutes that address this issue provide that in the event that a tenant's personal property does not sell at auction, the landlord may dispose of it.[126]

III. PROPERTY WRONGS OF SELF-STORAGE LAW

Part II of this Article brought to the attention of readers the widespread American cultural development of storing massive amounts of personal property in self-storage facilities across the United States.[127] The sheer volume of personal property now stored in self-storage facilities, measured in facility square footage and by the number of consumers using these facilities,[128] should be sufficient to create a public interest. If a similar quantity of real property and consumers were at stake in a more visible setting, such as with home foreclosures and the recent mortgage crisis,[129] no one would doubt that the public interest would be implicated.

---

120. VA. CODE ANN. § 55-419(E) (2007).

121. ALA. CODE § 8-15-34(13) (2002).

122. *See, e.g., id.*; COLO. REV. STAT. § 38-21.5-103(1)(j) (2010); CONN. GEN. STAT. § 42-164(d) (2008); DEL. CODE ANN. tit. 25, § 4904(h) (2009); FLA. STAT. § 83.806(8) (2004); LA. REV. STAT. ANN. § 9:4759(11) (Supp. 2011); OR. REV. STAT. § 87.691(7) (2009).

123. *See, e.g.*, HAW. REV. STAT. § 507-66(b) (2006); IOWA CODE § 578A.4(8) (2008); MICH. COMP. LAWS § 570.525(14) (2009); NEV. REV. STAT. § 108.477(5) (2009); R.I. GEN. LAWS § 34-42-4(i) (1995).

124. *See, e.g.*, GA. CODE ANN. § 10-4-213 (2009); IND. CODE ANN. § 26-3-8-15(c) (2005); KAN. STAT. ANN. § 58-817(d)(2) (2005); MO. REV. STAT. § 415.415(3) (Supp. 2011); NEB. REV. STAT. § 69-2308(4) (2010).

125. ARIZ. REV. STAT. § 33-1704(E)(6) (2007); 73 PA. CONS. STAT. § 1913 (2008).

126. *See, e.g.*, GA. CODE ANN. § 10-4-213 (2009); MISS. CODE ANN. § 85-7-125(e) (1999); OHIO REV. CODE ANN. § 5322.03(N) (2004); S.C. CODE ANN. § 39-20-45 (2010).

127. *See supra* notes 12–17 and accompanying text.

128. *See id.*

129. *See generally* Jeffrey D. Jones, *Property and Personhood Revisited*, 1 WAKE FOREST J.L. & POL'Y 93 (2011) (comparing demands for government intervention in the

Two other reasons weigh in favor of scrutinizing state-by-state practices of self-storage industry. First, personal property loss through self-storage default appears to be both a regular occurrence and a boom industry in periods of economic downturn.[130] Second, the procedures for handling the personal property of defaulted self-storage tenants are worrisome. In particular, the practice of unmonitored auctions of tenant property and the freedom of landlords in many states to sell tenant property by whole lot, without regard to the individual value of items, seems suspect.[131] Also, the opportunities for landlords to steal valuable tenant personal property prior to conducting public auctions seem abundant and undetectable.

This section discusses three property wrongs of self-storage law. First, self-storage statutes severely limit or wholly eliminate tenant remedies under tort law and the law of bailments, in effect giving self-storage landlords an absolute right of negligence with respect to tenant property. Second, the duty to conduct commercially reasonable sales appears to not be present in the context of self-storage auctions, and the absence of this duty results in windfalls to landlords and auctioneers that would otherwise be applied to tenant debt obligations. Third, some of the same public policy concerns that led to greater protection of residential property rights are present with regard to self-storage leases.

The section closes with a discussion of the contractual allocation of risk and the restrictions on what tenants may store in self-storage facilities. These restrictions commonly frustrate the purpose for which people rent self-storage units. Indeed, compliance with the contractual limitations of self-storage leases limits tenants to the storage of "crap" and expressly prohibits the storage of possessions having sentimental value or more than nominal economic value.[132] In other words, the common characteristic of many self-storage leases is that they contractually prohibit tenants from storing most of the items that would prompt rental of a self-storage unit in the first place.

A. Tort Law Misfires and the Elimination of Bailments

As explained in previous part of this article, self-storage landlords can only be held liable for damage or loss to tenant property under two circumstances: intentional or bad faith conduct by a landlord or its agents that causes damage to the tenant's property or[133] a violation of the express terms of state's self-storage facility act.[134] As demonstrated by the

---

mortgage and home foreclosure crisis to the absence of such demands for intervention in self-storage dispossession and auctions).

130. *See supra* note 18 and accompanying text.

131. *See supra* notes 102–13, 116–26 and accompanying text.

132. *See infra* Part III.D.

133. *See supra* Part II.B.

134. *See supra* Part II.B.

previously discussed cases, the statutory and contractual limitations on common law tort remedies give self-storage landlords a very broad right of negligence with regard to the treatment of tenant property.[135]

The right of negligence of self-storage landlords would not be absolute without also eliminating the law of bailments. Bailments, a creature of contract and property, arise upon the express or implied delivery of personal property by one party to be held in trust by another party.[136] The contractual aspect of bailments flows from the fact that bailment duties are premised upon express or implied agreements between the parties.[137] The property aspect of bailments flows from the fact that bailment duties depend on the bailee's—the party to whom personal property is delivered—lawful possession and control of the goods.[138] At early common law, the duty of care expected of bailees toward bailor property was "slight, ordinary, or great" depending on whether the bailment itself was for the sole benefit of the bailor, for mutual benefit of bailor and bailee, or for the sole benefit of the bailee.[139] In practice, however, a duty of reasonable care for all bailments has supplanted the tiered approach.[140]

The law of bailments differs from the common law of negligence in one crucial respect, which is of great significance to self-storage leases and potential landlord duties of care. A party who brings a claim of negligence against another for the loss or damage to personal property has the burden of proving negligent conduct by the latter.[141] By contrast, a bailee who loses or damages bailed property *is presumed* negligent, and the bailee must compensate the bailor if unable to rebut the presumption.[142]

---

135. *See supra* Part II.B.

136. *See, e.g.*,. Kurt Philip Autor, *Bailment Liability: Toward a Standard of Reasonable Care*, 61 S. CAL. L. REV. 2117, 2124–31 (discussing contract-based and property-based theories of bailment relationships).

137. *Id.*

138. *Id.*

139. *Id.* at 2131–33.

140. *See* RAY ANDREWS BROWN, THE LAW OF PERSONAL PROPERTY § 11.1 (3d ed. 1975) (noting that at early common law, "the bailee of goods was unconditionally and absolutely liable to return them to the bailor on demand," but that the tiered approach began with the English case, *Coggs v. Bernard*, 2 Ld Raym (KB) 909 (1703)). According to Brown, although the *Coggs* decision set in motion the diversification of bailee duties according to fault, the details of that "scheme of classification have not survived sufficiently to have direct importance in bailment law today." *Id.* Rather, the three-tiered benefit-based scheme of classification is owing to Justice Joseph Story. *Id.*

141. *See* R.H. Helmholz, *Bailment Theories and the Liability of Bailees: The Elusive Uniform Standard of Reasonable Care*, 41 U. KAN. L. REV. 97, 102 (1992) ("If a negligence standard is applied in cases where the goods have been lost or damaged, it will be incumbent upon the bailor to prove that the bailee's lack of ordinary care has caused the loss. The party alleging negligence, the bailor, must accept the burden of proof.").

142. *See* BROWN, *supra* note 140, § 11.8 ("It is a general rule that when the bailor has shown delivery to the bailee of the subject matter of the bailment and the latter has failed on demand to return the goods, or has returned them in a damaged condition, the bailor has

Most states' self-storage statutes make clear that self-storage alone cannot create a bailment relationship.[143] Georgia's Self-Storage Facility Act is typical.[144] It provides: "Except as otherwise specifically provided in this rental agreement, the exclusive care, custody, and control of any and all personal property stored in the leased space shall remain vested in the Occupant."[145] The inclusion of a statement disclaiming bailments in self-storage leases is also an industry best practice.[146] Together, these practices virtually eliminate the common law duty of care that normally arises when one person exercises control over another person's property.[147] Finally, to avoid circumstances that would otherwise give rise to a bailment relationship, the industry wisdom is, first, that facility owners require tenants to furnish their own locks,[148] and, second, that facility owners never enter tenant space or handle tenant property outside of the default process.[149]

---

made out a prima facie case for recovery of the value of the goods or for the damage inflicted upon them. He has satisfied the burden of going ahead with the evidence and is entitled to have the case left to the jury under instructions to find for the plaintiff if the jury believes that the facts are as testified to by the plaintiff. If, indeed, the testimony of the plaintiff is unimpeachable and of such character that no reasonable body of men could do otherwise than accept the same as true, the court should direct a verdict in favor of the plaintiff. It is customarily said in such cases that proof by the bailor of the delivery of the goods to the bailee and the latter's failure to return the same on demand free from damage is prima facie evidence that the loss or damage was due to the negligence of the bailee.").

143. *See, e.g.*, GA. CODE ANN. § 10-4-213 (2009); MD. CODE ANN., COM. LAW §18-505 (2005); MISS. CODE ANN. § 85-7-127 (2010) (stating that the exclusive care, custody and control of all personal property stored in the leased self-storage space remains vested in the occupant).

144. GA. CODE ANN. § 10-4-213 (2009).

145. *Id.*

146. *See* Jeffrey Greenberger, *Looming Legal Issues*, INSIDE SELF-STORAGE (Mar. 1, 2003), http://www.insideselfstorage.com/articles/2003/03/looming-legal-issues.aspx; Jeffrey Greenberger, *The Danger of Creating a Bailment*, INSIDE SELF-STORAGE (Oct. 1, 2004), http://www.insideselfstorage.com/articles/2004/10/the-danger-of-creating-a-bailment.aspx. Both web articles appear in the online publication, Inside Self-Storage ("ISS"), a leading trade publication for the self-storage industry. Additionally, Mr. Greenberger is an expert in the field of self-storage law and has accumulated a substantial number of online articles addressing legal issues in self-storage law. *See* SELF STORAGE LEGAL.COM, http://www.selfstoragelegal.com (last visited Sept. 1, 2011).

147. *See* Autor, *supra* note 136, at 2124–31.

148. *See* Jeffrey Greenberger, *Master Key Lock Systems for Your Self-Storage Facility*, INSIDE SELF-STORAGE (July 1, 2005), http://www.insideselfstorage.com/articles/2005/07/master-key-lock-systems.aspx; *Frequently Asked Questions: 10 Things You Need to Know About Getting into Self-Storage*, INSIDE SELF-STORAGE, http://www.insideselfstorage.com/faq.aspx (last visited Aug. 27 2011).

149. *See* Amy Brown, *The Risk of Retaining Tenant Keys*, INSIDE SELF-STORAGE (Dec. 1, 2002), http://www.insideselfstorage.com/articles/2002/12/the-risk-of-retaining-tenant-keys.aspx; Jeffrey Greenberger, *Accepting Deliveries for Customers*, INSIDE SELF-STORAGE

Another problem arises when self-storage landlords seize tenant property as part of the default process while expressly refusing tenant access. Clearly, under such circumstances a landlord holds exclusive custody and control of tenant property as would normally create some kind of bailment relationship. However, the statutory prohibition and contractual waiver on bailments prevent the creation of a bailment *even after* tenant default and the exercise of landlord lien rights when landlords unquestionably have lawful possession and exclusive control of tenant property for purposes of sale at auction.[150] Indeed, Georgia's Self-Storage Facility Act prevents the creation of a bailment in this situation:

> "Except as otherwise specifically provided in this rental agreement, the exclusive care, custody, and control of any and all personal property stored in the leased space shall remain vested in the Occupant. *The Owner does not become a bailee of the Occupant's personal property by the enforcement of the Owner's lien.*"[151]

Thus, self-storage landlords have their cake and eat it too. State law frees self-storage landlords from the limits placed upon traditional creditor-debtor remedies and authorizes them to place liens on stored tenant property to satisfy debt,[152] yet upon exercising the lien power by seizing and selling tenant property, self-storage landlords are spared the common law duty of reasonable care that typically comes with gaining lawful possession of another's property for mutual benefit.[153] Self-storage landlords are not required to exercise reasonable care even though the default and auction periods present the greatest likelihood of mistreatment of tenant property. During these times, tenants' rights are at their weakest, and law enforcement's ability to timely detect and prevent misconduct is at its lowest.

I propose two policy explanations for granting self-storage landlords such remarkable powers in contravention of the common law of property. Both rationales stem from a view of self-storage leases as typical commercial transactions. The first rationale supposes that the principles from the Uniform Commercial Code governing commercially reasonable sales should apply to self-storage leases just as they apply to any other commercial transaction. The second rationale supposes that self-storage leases are standard commercial leases devoid of the public policy concerns that have led to important residential consumer protection doctrines created by common law courts.[154]

(Nov. 1, 2002), http://www.insideselfstorage.com/articles/2002/11/accepting-deliveries-for-customers.aspx.

150. *See* statutes cited *supra* note 143 and accompanying text.

151. GA. CODE ANN. § 10-4-213 (2009) (emphasis added).

152. *See supra* Part II.A.

153. *See supra* notes 43–45.

154. *See infra* note 190 and accompanying text.

These rationales are questionable on the ground that self-storage leases are not typical commercial transactions. First, a sale or lease that is commercially reasonable between two companies may not be so between a company and a consumer, and self-storage leases fall into the latter category. This premise is well-grounded in U.S. consumer protection law but appears absent in the context of self-storage leases. Second, self-storage leases contain characteristics of both commercial and residential leases, although the latter features continue to be ignored by courts.[155] This suggests that courts have made a policy choice about how to characterize self-storage leases that is open to question.

B. Bona Fide Purchases and the Price Function in Commercially Reasonable Sales

The majority of state self-storage statutes follow a procedural approach to a landlord's duty to conduct a "commercially reasonable sale" of tenant property following default.[156] That approach presumes that compliance with the procedures for sale set by the state self-storage statute satisfies every duty of a commercially reasonable sale.[157] Put another way, under the procedures-based approach to commercial reasonableness, price is not considered an "aspect" or "term" of commercial reasonableness.[158] The procedures-based approach to commercially reasonable sales prevents a tenant from asserting that a sale was commercially unreasonable because of a low selling price.[159] In jurisdictions that use a procedural approach, a tenant cannot make the case—other than to show that statutory sales procedures were violated—that a landlord's sale of the tenant's property should have yielded greater proceeds to offset the default balance.[160]

New York's law makes this explicit:

> 8. Pricing. The fact that a better price could have been obtained by a sale at a different time or in a different method from that selected by the owner is not of itself sufficient to establish that the sale was not made in a commercially reasonable manner. If the owner either sells the goods in the usual manner in any recognized market therefor, or if he sells at the price current in such market at the time of his sale, or if he has otherwise sold in conformity with commercially reasonable practices among dealers in the

155. *See id.*

156. *See, e.g.*, CAL. BUS. & PROF. CODE § 21707 (2011); FLA. STAT. § 83.806 (2004); NEV. REV. STAT. § 108.477 (2009); N.Y. LIEN LAW § 182 (McKinney Supp. 2011).

157. *See supra* note 113 and accompanying text.

158. Michael Korybut, *Searching for Commercial Reasonableness Under the Revised Article 9*, 87 IOWA L. REV. 1383, 1386 (2002).

159. *See* N.Y. LIEN LAW § 182(8) (McKinney Supp. 2011); WYO. STAT. ANN. § 29-7-105(k) (2011).

160. *See id.*

> type of goods sold, he has sold in a commercially reasonable manner. A sale of more goods than apparently necessary to be offered to insure satisfaction of the obligation is not commercially reasonable except in cases covered by the preceding sentence.[161]

Self-storage statutes also permit a defaulted tenant's property to be sold "singly, in lots or as a whole."[162] It is difficult to see how selling a tenant's property as a whole, which is "the usual manner" among self-storage facility operators, is consistent with any duty of commercially reasonable sale that has teeth. Suppose Tenant has defaulted on a self-storage agreement and Landlord has denied Tenant access, seized the unit, and scheduled a public auction. Tenant owes $300 in arrears to recover the unit before auction. Tenant's unit has property with a combined fair market value of $600, all contained within seven sealed boxes. Landlord opts to sell the lot as a whole, under terms similar to the auction described in the introduction to Part II of this Article.[163] That is, bidders are not allowed to enter the unit, much less view the particular items inside of the boxes. Instead, bidders are permitted only to shine their flashlights inside the unit, over and between the boxes, and then use their imaginations to determine how much to bid.[164]

Under these auction conditions, no bid is likely to approach the $600 value of Tenant's property or even the value the items would be estimated to have if the items inside the boxes were placed on display for individual valuation. Any shortfall in proceeds from such a sale harms Tenant. If Landlord could have recouped $400 by placing the items on display, Tenant would have been entitled to $100. Instead, if Landlord only recoups $150 through a sale of the storage unit as a whole, Tenant is left owing $150. However, the procedures-based approach of evaluating whether a self-storage auction is a commercially reasonable sale precludes arguing that this sort of sale is commercially unreasonable because of a low sale price.[165]

The procedures-based approach to commercially reasonable sales in the self-storage context is not new. Instead, the procedures-based approach appears to be modeled after the majority approach to commercially reasonable sales followed by the Uniform Commercial Code ("U.C.C.) and the Revised U.C.C. This is true even though the U.C.C. does not apply to self-storage agreements unless, in conjunction with such agreements, "an owner issues a warehouse receipt, bill of lading or other document of title for the personal property stored . . . ."[166]

---

161. *See, e.g.*, N.Y. LIEN LAW § 182(8) (McKinney Supp. 2011).

162. *See supra* note 116 and accompanying text.

163. *See* text accompanying *supra* note 11.

164. *Id.*

165. *See supra* note 159 and accompanying text.

166. *See* COLO. REV. STAT. § 38-21.5-101(7) (2010); *see also* CAL. BUS. & PROF. CODE § 21701(a) (West 2003).

The drafters of the Revised Article 9 considered two approaches—procedures-based and proceeds-based—to the duty of commercially reasonable sales.[167] The foundational question was, "should solely the . . . sale process and its procedural regularity measure the sale's commercial reasonableness or should the main focus or inquiry be the proceeds produced by the sale?"[168] Under a proceeds test, the primary concern of a court's commercial reasonableness review was the reasonableness of the sale's proceeds rather than the reasonableness of its procedures.[169] Unless the secured party proves that a low sale price for collateral was justified, a low sale price alone could render an otherwise procedurally regular sale commercially unreasonable.[170] Under a procedures test, the regularity and reasonableness of a sale's procedures takes priority in a court's analysis.[171] In such jurisdictions, a low price alone cannot support a claim for commercial unreasonableness.[172]

Revised Article 9 does not expressly require one test or the other.[173] The ascendant interpretation and practice, however, is to follow the procedures test, under which "commercial reasonableness should be measured primarily through examination of the sale's procedures rather than its price."[174] One great virtue of the procedures test is that it spares courts and litigants of "valuation battles" and the need to consider secondary source evidence of value and fair price.[175] When the revised U.C.C. sale procedures have been followed, courts must conclude that a fair market price has been reached.[176] Exported into self-storage statutes, the procedures test for a commercially reasonable sale sanctions what has been described earlier: a defaulted tenant's belongings may be sold for $1, even if otherwise intuitive measures requiring little or no cost—such as unboxing items for bidders to view—would have yielded much greater proceeds.[177]

There are two problems with the procedures test for commercial reasonableness as it currently exists in state self-storage statutes. First, the bar for procedures simply is too low. In the U.C.C. context, secured parties must "use the market sale practices that she in good faith believe[s are] best suited to maximize the collateral's price and which are reasonably available to the secured party."[178] The rationale of the U.C.C. procedures test is "that only through exposure to the marketplace could a collateral's fair market

167. *See* Korybut, *supra* note 158, at 1386.
168. *Id.*
169. *Id.* at 1387.
170. *Id.*
171. *Id.* at 1386.
172. *Id.* at 1387.
173. *Id.* at 1432.
174. *Id.*
175. *Id.* at 1387.
176. *Id.*
177. *See* Mesh, *supra* note 11.
178. Korybut, *supra* note 158, at 1392.

price be accurately determined for purposes of assessing a sale's commercial reasonableness and calculating a deficiency."[179] It is obvious that bidding on cardboard boxes with the collateral inside—or layered in a self-storage unit so that only items in front are visible—is unsatisfactory "exposure" to any marketplace.

Second, even though the revised U.C.C. practice favors the procedures test, price continues to have a role in courts' assessment of commercial reasonableness.[180] The Official Comments to revised Article 9 state that, "[w]hile not itself sufficient to establish a violation of this Part, a low price suggests that a court should scrutinize carefully all aspects of a disposition to ensure that each aspect was commercially reasonable."[181] Thus, *in U.C.C. practice*, although a low price alone cannot render a sale commercially unreasonable, it is a signal to courts to carefully scrutinize every other aspect of the transaction.[182] Despite regular occurrence of disturbingly low sale prices at self-storage auctions, courts refuse to apply any judicial scrutiny to such auctions.[183] If, in the self-storage context, the duty to hold a commercially reasonable sale does not require self-storage landlords to make collateral visible to bidders, then the bar for commercial reasonableness—just in terms of procedures—is stunningly low.

C. A Parallel to Residential Landlord-Tenant Law

Personal property is everything one owns that is not real property.[184] Real property admits of a further policy distinction between residential and commercial property.[185] To designate real property as "residential" has, in our society, the legal effect of strengthening owners' property rights based upon a variety of public policy considerations, including the need for affordable housing,[186] the importance of shelter to personal welfare,[187] and the special place in the American imagination held by the home.[188] Likewise, to designate real property as "commercial" signals to market participants that government has no special interest in the kinds of dealings

---

179. *Id.* at 1386–87.

180. *Id.* at 1433.

181. *Id.*

182. *Id.*

183. *See* Kane v. U-Haul Int'l Inc., 218 F. App'x 163 (3d Cir. 2007); Mesh, *supra* note 11.

184. BLACK'S LAW DICTIONARY 1337 (9th ed. 2009).

185. *Id.*

186. *See generally* Stephanie M. Stern, *Residential Protectionism and the Legal Mythology of Home*, 107 MICH. L. REV. 1093 (2009) (discussing the many protections afforded to residential housing).

187. *See* Jones, *supra* note 129, at 127.

188. *See generally* D. Benjamin Barros, *Home as a Legal Concept*, 46 SANTA CLARA L. REV. 255 (2006) (discussing the favorable treatment of homes in various legal contexts and whether or not this treatment is justified).

that occur in conjunction with the property beyond ordinary above-board contracting.[189]

As demonstrated by Judge Fisher's opinion in *Kane*, courts routinely treat self-storage cases as purely commercial disputes.[190] There is an explanation for this, however. Self-storage facilities are commercial enterprises. In fact, the real property purchased or leased by facility owners is usually acquired from other commercial enterprises. Most importantly, no one lives, or is legally permitted to live, in a self-storage unit.[191]

However, the self-storage unit—the external shelter for personal property—is not the only property involved in self-storage leases that might influence what the law should be. The personal property placed within the unit may itself be considered residential or commercial in character. Attending to the character of personal property stored allows a new distinction not yet addressed by courts. The shift in attention distinguishes warehouses located at docks, shipyards, railroads, airlines, and industrial farms, all of which lease space for storage of bulk goods or other commercial products, from self-storage facilities that lease space for the storage of individual personal property.[192]

---

189. *See* Javins v. First Nat'l Realty Corp., 428 F.2d 1071, 1080 (D.C. Cir. 1970) (creating an implied warranty of habitability for residential housing and explaining the policy bases for treating residential leases differently than commercial leases).

190. *See* Kane v. U-Haul Int'l Inc., 218 F. App'x 163 (3d Cir. 2007). In affirming summary judgment for the defendant, Judge Fisher reasoned that a self-storage contract is more like a commercial lease than a residential lease:

> The contract for the storage units clearly was standardized. However, it cannot be said that the Appellants had no opportunity to make any choices. They were provided with the option of purchasing insurance to protect against negligence for an additional fee. The public interest is not affected in light of the fact that the opportunity to elect insurance for an additional reasonable fee existed. A contract for self-storage cannot be equated with a residential lease. The prohibition of enforcing exculpatory clauses in residential leases is based on housing shortages, especially affordable housing, the need for which has been recognized by the New Jersey legislature. Additionally, the exculpatory clause and offer of insurance were both clear in the contracts signed by the Appellants. A self-storage contract is more akin to a lease for commercial space. Therefore, we agree with the District Court's determination that no unequal bargaining power existed that would make the exculpatory clause unenforceable.

Id. at 166.

191. *See, e.g.*, ALA. CODE § 8-15-31(8) (2002) ("No occupant shall use a self-service storage facility for residential purposes."); ARIZ. REV. STAT. ANN. § 33-1702(B) (2007) ("An occupant shall not use a leased space for residential purposes."); ARK. CODE ANN. § 18-16-403(a)–(b) (2003) ("An operator may not knowingly permit a leased space at a self-service storage facility to be used for residential purposes. . . . An occupant may not use a leased space for residential purposes.").

192. *See* 78 Am. Jur. 20 *Warehouses* § 2.

Furthermore, warehouses store commercial personal property, which is typically held in the name of a corporation or other business entity, for the purpose of investment and commercial gain.[193] By contrast, the property stored at self-storage facilities typically is residential personal property purchased by tenants and held for personal enjoyment as part of a fully functioning home and life.[194] Even though state courts have not recognized this distinction, nearly every state legislature has through the creation of self-storage facility acts that are separate from commercial warehousing laws.[195]

A ready objection to the distinction between residential and commercial personal property supposes that personal property placed in self-storage is, by definition, not residential because such property is not kept at a home residence. But observe how quickly this argument falls apart. Home furniture, tax records, family photographs, heirlooms, etc., do not cease to be residential just because they are stored off-premises when not in use. The aforementioned personal property is residential whether it is kept in one's basement, one's garage, *or one's self-storage unit.* Similarly, an individual who sets up a corporation and purchases supplies and products in the corporation's name has purchased commercial personal property, even if that personal property is stored at home and the business is operated from home.

The error of the foregoing objection is that the locus of personal property is not a sound indicator of its status as residential or commercial. Instead, as with real property, the distinction between residential and commercial property turns on usage. For example, an individual may lease self-storage space in order to store products used in operating a home mail order business or in the production of illegal drugs. In these situations, the argument that such personal property is residential fails because the individuals are storing products used for investment and business purposes, much like the railroads, airlines, and industrial farms. One might still conclude that residential personal property placed in self-storage facilities deserves less protection than property stored at a personal residence, but the basis for differential treatment should not be that such property is "commercial" by virtue of its storage away from home.

If the distinction between residential and commercial personal property is accepted, it is no longer doctrinally obvious or inevitable that self-storage leases are commercial leases; rather, a judicial policy choice has been made about self-storage leases that is open to question.

---

193. *See id.*

194. *See* Mesh, *supra* note 11.

195. For example, Arizona's commercial warehousing laws are codified under its section of the U.C.C., ARIZ. REV. STAT. ANN. tit. 47, ch. 7 (2007), but its self-storage provisions are codified under the general provisions on property, ARIZ. REV. STAT. ANN. tit. 33, ch. 15 (2007).

Even if personal property placed in self-storage facilities is residential rather than commercial in character, Judge Fisher's opinion in *Kane* suggests that the public interest is not so affected as to warrant judicial scrutiny of self-storage agreements. However, several factors discussed earlier in this Article undercut this claim:

- The boom in the self-storage industry is evidence of a shortage of home space for personal property.[196]
- Where personal property has welfare or identity functions—such as clothing necessary for work or medical devices necessary for health—the public interest is affected.[197]

Particularly in cases where property placed in self-storage is the result of economic hardship, such as divorce, homelessness, or illness, there is a public interest in ensuring that auctions conform with something similar to due process. By due process, I am not referring to the Constitutional Due Process protection bracketed in the introduction, but rather to the legal process which the common law of property has long required prior to dispossessing people of what they own.

Finally, Judge Fisher's suggestion of equal bargaining power between self-storage tenants and landlords rings hollow. Self-storage tenants do not have their own lobby. Self-storage contracts, which place virtually all risk upon tenants, are not negotiable.[198] And insurance—which exists primarily to protect landlords against their own misconduct rather than the tenant's own behavior—addresses little that is askance in self-storage law.[199]

D. The "Crap" Rule

U.S. property law carries the doctrinal presumption that all real property is unique in character.[200] This presumption necessitates the remedy of specific performance in real property transaction disputes.[201] In other words, in cases where a seller of real property seeks to escape from a contract for sale, the presumption renders expectation damages inadequate

196. *See supra* notes 12–17 and accompanying text.

197. *See supra* note 88 and accompanying text.

198. *See If You Rent a Storage Facility, What Kinds of Contractual Rights and Obligations Do You Have?*, HOMELESS LAW BLOG (Jan. 31, 2009, 10:19 PM), http://homelesslaw.wordpress.com/2009/01/31/if-you-rent-a-storage-facility-what-kinds-of-contractual-rights-and-obligations-do-you-have/.

199. *See supra* Part II.B.

200. *See* Tanya D. Marsh, *Sometimes Blackacre Is a Widget: Rethinking Commercial Real Estate Contract Remedies*, 88 NEB. L. REV. 635, 636–48 (2010) (discussing the "uniqueness" doctrine in real property law).

201. *Id.* at 636.

to make the would-be buyer whole.[202] Beyond this transactional presumption, scholars have proposed additional doctrines that account for circumstances where real property is or should be granted special legal protection. These doctrines fall mainly into two categories. One category marks real property as special according to its connection with individual or group *identity*, such as cultural property.[203] The other category marks real property as special according to its connection with individual or group *welfare*, such as residential housing.[204]

Personal property can be special according to its connection with identity or welfare in the same way that real property can. For example, property such as family heirlooms, personal records, collectors' items, the meager belongings of homeless individuals, and the residue of possessions owned by evicted residential tenants or foreclosed-upon home owners all have similar connections with their individual owner's identity and welfare that cause the property to be unique.

Most state self-storage statutes offer no special protection whatsoever for personal property with welfare functions. Indeed, few state laws require facility owners to grant defaulted tenants access to stored property necessary for personal welfare, such as personal records, health aids, or clothing.[205] Furthermore, self-storage leases routinely require that tenants agree not to store personal property with "sentimental value"—that is, personal property with identity functions.[206]

When the lack of special access rights to personal property with welfare functions is paired with the prohibition on storing personal property with identity functions, the business model of the self-storage industry becomes clear. Its sole product is the storage of *"crap"*—personal property that tenants neither highly value nor need—*and only "crap."* Tenants who know their rights would be foolish to store any personal property with welfare functions and would be in breach of contract for storing personal property with identity functions.[207]

At first glance, the service of storing relatively valueless things in exchange for money seems unproblematic. In this instance, however, social

---

202. *Id.*

203. *See generally* Patty Gerstenblith, *Identity and Cultural Property: the Protection of Cultural Property in the United States*, 75 B.U. L. REV. 559 (1995) (arguing that society as a whole, not just the government, must take steps to ensure that cultural property is protected).

204. *See* Stern, *supra* note 186, at 1099–1105 (discussing the many protections afforded to residential housing).

205. *See supra* note 88 and accompanying text.

206. *See* Public Storage, *Lease/Rental Agreement* 3 (Aug. 18, 2010) (on file with the Tennessee Law Review).

207. *See* Low v. Penn Self Storage, No. 2010AP132, 2010 WL 5186050, at *2 (Wis. App. Dec. 23, 2010) (upholding a provision in a self-storage lease, in which the lessee agreed "**not to store property with a total value in excess of $15,000.00**" or "irreplaceable property such as books, writings, objects which have an unknown immediate resale market value, or objects which have a special or emotional value to Lessee.").

tendencies create complications. The tenant's duty to store only low-value personal property is counter to reasonable tenant expectations and will often prove to be a difficult duty to meet. When people rent self-storage units, they do so to store pre-identified things that are personally valued above the monthly costs of storage and the associated risks. The contractual duty to store only low-value personal property would, in many cases, require would-be renters to forego self-storage rental altogether or replace preselected storage items of personal value with "crap" that complies with the self-storage contract limitations. In fact, under many self-storage agreements, tenants would be in breach for storing reasonable and expected items such as family Christmas ornaments or inherited china used only during holidays.[208] Similarly, a person would likely be in breach for storing the balance of what once fit in a large home, but which does not fit in a small apartment.[209] Additionally, the prohibition on storing much-valued personal property could be considered unreasonable for the homeless or working poor, residential evictees, persons in home foreclosure, or individuals who default because of bankruptcy, medical emergency, or military leave.

The central policy concern raised by what may be called the "crap" rule is not protection of the least well off; rather, the central policy concern is two-fold. First, the low-value personal property restrictions imposed upon tenants by self-storage agreements are unreasonable when viewed in light of the reasons for which tenants rent self-storage in the first place. Second, to the extent that some personal property is entitled to special legal protections, such protections may extend to personal property placed in self-storage.

IV. CONCLUSION: PERSONAL ACCUMULATION AND THE PROPERTY ETHIC "POSSESS LESS"

My goal in writing this Article was to reveal the current state of the self-storage industry as a consumer protection problem. I do not argue that self-storage owners are a bad lot. Instead, I argue that self-storage facility owners enjoy legal powers that do not make sense from the perspective of property law. The perversion of property law present in self-storage facility acts, along with the imbalance of power between self-storage landlords and tenants, is explained by politics. The Self Storage Association is a powerful, well-organized lobby operating in most states.[210] State self-storage facility acts reflect this fact. Individual tenants lack the organization, the power, and often the wherewithal to detect violations of the limited rights that they do have.[211]

---

208. *See id.*

209. *See id.*

210. *See supra* note 12 and accompanying text.

211. In discussing the top five legal threats to self-storage operators in 2011, attorney

Reforming the practices of self-storage industry is easy enough. Simply reinstitute the common law of negligence and bailments, and make federal a distrait process similar to those already present in a few states. Regarding the valuation of personal property sent to auction, defaulted self-storage tenants should have protections similar to the debtor protections afforded to a defaulted mortgagor whose real property is auctioned by the mortgagee at a foreclosure sale. In the real property foreclosure context, many states have adopted measures to prevent the low bids that currently plague self-storage auctions.[212] Some of the following measures available in the real property foreclosure context could also be used to protect defaulting self-storage tenants: a statutory right of redemption, which allows a mortgagor to buy back the property for the price bid at the foreclosure sale for a designated period after foreclosure; the prohibition on deficiency judgments, which decreases a mortgagee's incentive to bid below the fair market value of the

---

Jeffrey Greenberger, a lead industry attorney, points to the television program *Storage Wars*:

> As an [sic] lawyer, I see errors made in these sales, and it drives me crazy. I keep wondering, when is a tenant going to watch his own goods being sold on national television, recognize the facility did something in error, and file a lawsuit for violations, using the taped raw or edited footage of the sale as proof positive? That is perhaps the smaller of my two concerns.
>
> My larger concern is judges and potential jury members, as well as self-storage tenants, are beginning to believe every self-storage unit is a treasure trove for which you'll get so much more than what you paid (i.e., the operator never sells a unit for its full value). . . .
>
> Let's consider a theoretical wrongful sale at a self-storage facility. The tenants ("victims") sue and allege all sorts of missing valuable property. (For the purposes of this example, you can substitute for wrongful sale other incidents such as theft while your overlock is on the unit, building fires, wrongful access, etc.). In the old days, we would at least have an argument that a tenant who was habitually late in paying his $75 a month rent probably didn't have $100,000 worth of antique guns in his unit. Judges simply didn't believe that people who struggle to pay their bills every month would have such valuable property, with the exception of emotional and sentimentally important items.
>
> Now judges and juries are going to think differently. We're going to allege that what was [sold] at lien sale or what was in the unit when we inventoried it and put our overlock on it was a mattress, box spring and some old clothing. Tenants are going to allege items such as fishing poles, antique guns and commemorative coins, all of which have been featured on "Storage Wars." From now on, we're going to be in a battle to disprove the value of every item instead of trying to make the tenants prove there was value to their items.

Jeffrey Greenberger, *The Top Five Legal Threats for All Self-Storage Operators in 2011*, INSIDE SELF-STORAGE (Jan. 20, 2011), http://www.insideselfstorage.com/articles/2011/01/the-top-five-legal-threats-for-all-self-storage-operators-in-2011.aspx?pg=3.

212. *See* JOSEPH WILLIAM SINGER, PROPERTY LAW: RULES, POLICIES, AND PRACTICES 896 (5th ed. 2010).

property; a mortgagor's ability to bring unjust enrichment claims against mortgagees in cases where a mortgagee buys the property at a low price and resells it within a short period of time for a windfall; and judicially supervised sales in which courts ensure that the sale price is adequate.[213] Alternatively or additionally, the accumulation behavior of would-be renters could be altered by requiring mandatory disclosure of tenant inventories and of how few tenant rights exist after personal property is stored.

There is good indication that the practices of the self-storage industry could also be changed if the issue of commercially reasonable sales reached juries. In *Cook v. Public Storage, Inc.*, a 2008 case, plaintiff Cook sued Public Storage, Inc. after the company auctioned off personal property of Cook's that was stored in one of Public Storage's rental units.[214] Cook sued Public Storage for failure to provide proper notice of the sale (the two notices that Public Storage mailed were likely misaddressed and were returned undeliverable) and for failure to conduct the sale in a commercially reasonable manner in violation of Wisconsin's self-storage facility act.[215] The winning bid for Cook's property at auction was $660.[216] The jury determined that a commercially reasonable sale of Cook's property would have yielded $5000 (coincidentally, the contractual limit of Public Storage's liability under the rental agreement).[217] More important, the jury awarded Cook $19,000 in compensatory damages and $100,000 in punitive damages.[218] In addition, the court awarded Cook's counsel $262,500 in attorney's fees plus $19,654.02 in costs and disbursements.[219] The Wisconsin Court of Appeals affirmed, and the Wisconsin Supreme Court denied Public Storage's petition for review.[220]

Victories such as Cook's are very rare in self-storage legal disputes. In most states, the satisfaction of minimal procedural requirements for auctions, such as proper pre-sale notice, immunizes facility owners. But this is not enough for the self-storage industry. Currently, state self-storage associations around the country are pushing legislation to eliminate what few consumer protections remain for self-storage renters.[221] In coordination with the Self Storage Association, state self-storage associations are mounting concerted efforts to lower their own costs and further decrease potential liability by:

---

213. *See id.* at 895–96.

214. *Cook v. Public Storage, Inc.*, 761 N.W.2d 645, 649–503 (Wis. Ct. App. 2008).

215. *Id.* at 651.

216. *Id.*

217. *Id.* at 652.

218. *Id.*

219. *Id.*

220. *Id.*; *Cook v. Public Storage, Inc.*, 759 N.W.2d 772 (Wis. 2008).

221. For the following discussion, I am indebted to several industry experts who preferred not to be cited in this article.

- Eliminating the requirement that notice of auctions be published in local newspapers in the jurisdiction where the self-storage facility is located and replacing it with notice on publicly available websites;[222]
- Eliminating the requirement that renters be notified of the auction of their stored property via certified and registered mail and allowing facility owners to give notice through regular mail with proof of mailing;[223]
- Eliminating the mailing requirement altogether and allowing facility owners to notify tenants of default and sale via e-mail;[224]
- Further limiting facility owners' contractual liability by amending state self-storage facility acts to include a provision stating that if the rental agreement contains a limit on the value of property stored in the occupant's storage space, the limit shall be deemed to be the maximum value of the property storage in that space.[225] This provision gives facility owners greater discretion and a permanent and powerful incentive to always undervalue tenant property;
- Where the duty to return personal papers, photographs, etc. exists, transferring the duty from facility owners to auction buyers;[226]

---

222. *See, e.g.*, MICH. COMP. LAWS § 570.525(5) ("[A]n advertisement of the sale or other disposition shall be published once a week for 2 consecutive weeks in the print *or electronic version* of a newspaper of general circulation in the area where the self-service storage facility or self-contained storage unit is located or posted once per week for 2 consecutive weeks *on a publicly available website identified in the rental agreement.*" (emphasis added)).

223. *See, e.g.*, MICH. COMP. LAWS § 570.525(3) ("A notice given pursuant to this section shall be presumed delivered when it is deposited with the United States postal service and properly addressed with postage prepaid or when it is transmitted by electronic mail to the tenant's last known electronic mail address. An owner who gives notice under subsection (2) shall make an affidavit stating how and when the notice was delivered to the tenant and shall attach a copy of the notice to the affidavit. The owner shall retain the affidavit for introduction into evidence in any potential action . . . .").

224. *See, e.g.*, ME. REV. STAT. ANN. tit. 10 § 1375(2) (2011) ("As soon as the occupant is in default and before conducting a sale under subsection 1, the operator shall: A. Send a notice of default *by verified mail and by either first-class mail or electronic mail* to the occupant at the occupant's last known address or other address set forth by the occupant in the rental agreement . . . ." (emphasis added)).

225. *See, e.g.*, ME. REV. STAT. ANN. tit. 10 § 1375(2) (2011) (***"Value of stored property.*** If a rental agreement contains a limit on the value of personal property that may be stored in the occupant's leased space, the limit is deemed to be the maximum value of the stored personal property and the maximum liability of the operator for any claim.").

226. For example, WASH. REV. CODE ANN. § 19.150.070 (West 2011) prohibits facility owners from selling stored personal papers and photographs at auction and requires facility owners to keep personal papers and photographs for six months. After six months, facility owners may dispose of the property in a reasonable manner. The most obvious interpretation

- Adding to self-storage facility acts a definition of "abandoned leased space" that would allow facility owners to immediately seize and dispose of tenant property whenever such property is found unlocked and contains personal property below a certain value (*as adjudged by the facility owner*).[227]

In light of these ongoing developments, a federal consumer protection statute seems in order.

Despite the self-storage industry's lack of concern for tenants' property rights, something still may grate on the American psyche about helping self-storage tenants. Some may feel that the common law should not come to the rescue of self-storage tenants. After all, what is preventing distressed tenants from gathering up their goods before they default? Given much of the "crap" that is found in self-storage facilities, landlords often prefer this practice to having to conduct auctions because the unit is immediately ready to re-lease. Furthermore, some might say that personal accumulation in America has gone too far[228] and that dispossessed tenants may be left better off and a bit morally corrected by the whole affair.

On the subject of self-storage, American cultural mores may be at odds with the common law of property. Where the common law of property sees corporate exploitation of a relatively weak consumer population, American cultural mores may see proof of a gluttonous American population. Criticism of the American penchant for accumulating "crap" is rampant.[229]

---

of this provision imposes a duty upon facility owners to search units and to remove and store personal papers and photographs before auctioning the unit. However, the industry appears to interpret "owner" as referring to buyers who obtain the personal property at auction. This interpretation transfers the duty to keep the personal papers and photographs of dispossessed tenants for six months from facility owners to buyers. (This observation is based on my own discussions with industry experts in researching this subject.).

227. *See, e.g.*, ME. REV. STAT. ANN. tit. 10 § 1378 (2011) ("In the case of an abandoned leased space, the operator has the right to immediately take possession of the leased space and dispose of any personal property in the leased space by any means at the operator's discretion.") "Abandoned leased space" is defined as "a leased space that the operator finds unlocked and empty or unlocked and containing personal property with a value less than $750 or a leased space possession of and all rights to which and any personal property within which have been surrendered to the operator by the occupant." ME. REV. STAT. ANN. tit. 10 § 1372(1-A) (2011).

228. *See supra* note 215 and accompanying text.

229. *See, e.g.*, Russell W. Belk et al., *Dirty Little Secret: Home Chaos and Professional Organizers*, 10 CONSUMPTION MARKETS & CULTURE 133 (2007) (exploring "the deep meanings of clutter" and "visualiz[ing] the issues in disorganization and frustrations of home clutter and chaos as well as the methods and results of [professional] organizers"); Kieran Doherty, *Clean Your Closet, Clear Your Mind*, ORGANIC STYLE, Oct. 2005, at 35; Nancy Keates, *The Struggle to Contain Ourselves*, WALL ST. J., Jan. 4, 2008, at W1; Lisa McLaughlin, *How to Live with Just 100 Things*, TIME, June 16, 2008, at 57 (chronicling a movement in the United States for every American to get rid of all but 100 things they own);

However, growth of personal property holdings is a characteristic of all affluent societies. More importantly, self-storage laws should not provide an opportunity to enact moral legislation designed to encourage us to make better property choices. Self-storage defaults already kick individuals when they are down. Using legislation to limit property rights in the self-storage default context would add moral insult to economic injury and seems to be an unduly harsh response that would accomplish little.

Nancy Stedman, *20 Minutes to a Clutter-Free House: Quick Easy Strategies that Really Work*, PREVENTION, Oct. 2004, at 133; Lauren Baier Kim, *Do We Need Bigger Homes or Less Stuff?*, WALL ST. J. BLOG (Jan. 4, 2008, 1:32 PM), http://blogs.wsj.com/developments/2008/01/04/do-we-need-bigger-homes-or-less-stuff/.

JUSTIFYING GINA

ABIGAIL LAUREN PERDUE*

I. INTRODUCTION 1052
II. GENETICS 101 1055
A. *The Evolution of Genetics* 1055
B. *Decoding the Genetic Code* 1055
C. *The Human Genome Project* 1057
D. *From Pea Plants to Punnett Squares: Understanding Inheritance* 1058
E. *Mutation and Codominance* 1059
F. *Sex Linkage* 1059
G. *Heritability* 1060
H. *Ecogenetics* 1061
I. *Dominance or Destiny* 1062
III. EXPLORING TITLE II 1063
A. *Justifying GINA* 1074
B. *Is GINA Necessary?* 1074
C. *Did Pre-Existing Law Adequately Protect Against Genetic Discrimination?* 1078
D. *ADA* 1079
E. *EEOC Guidelines* 1080
F. *Pre- and Post-Employment Genetic Testing* 1081
G. *Defining "Disability" After Bragdon* 1082
H. *Regarded As "Genetically Disabled"* 1088
I. *Direct Threat to Self Defense* 1089
J. *Title VII* 1092
K. *OSHA* 1093
L. *Executive Order No. 13145* 1095
M. *State Legislation* 1095
N. *Genetic Exceptionalism* 1097
IV. AMERICA'S BRAVE, NEW WORKPLACE: COMPETING

---

* Abigail Lauren Perdue is an employment attorney who formerly taught employment discrimination law and Law, Genetics, and Policy at Washington and Lee University. She dedicates this Article to her parents, Janet and David Perdue, sister and nephew, Rachel and Heath Turner, her grandparents, Mary and Hoover Asbury and Maxine and Herbert Perdue, and her best friend, Amber Thompson Foster, as well as the rest of her family and friends whose unfailing support and love made it possible. She also wishes to thank Professor Kim Forde-Mazrui for his incredible guidance and insight throughout the drafting process. The views expressed herein are solely the views of the author writing in her personal capacity.

CONCERNS AND INTERESTS 1100
A. *Employer Concerns and Interests* 1100
B. *Employee Concerns and Interests* 1106
V. LEGISLATIVE SPLICING: IMPROVING TITLE II OF GINA 1111
A. *Expand the Definitions of "Genetic Information" and "Genetic Test"* 1111
B. *Rephrase Section 2(3)* 1112
C. *Clarify Prohibited Conduct* 1112
D. *Improve Confidentiality by Sealing Records in GINA Cases* 1112
E. *Clarify GINA's Relationship to Other Federal Legislation* .. 1112
F. *Clarify Whether GINA Applies Extraterritorially* 1113
G. *Include a More Comprehensive Sunset Provision* 1113
H. *Adopt an Informed Consent Model* 1113
I. *Clarify GINA's Impact on Workers' Compensation* 1114
J. *Add Lab Restriction* 1114
K. *Consider Allowing Disparate Impact Claims* 1114
L. *Consider Adding a BFOQ Provision* 1115
M. *Clarify the Applicability of the Direct-Threat-to-Self Exception* 1115
N. *Address Genetic Propensity for Violence* 1115
VI. CONCLUSION 1117

I. INTRODUCTION

In the nineteenth century, Gregor Mendel performed pea plant experiments in the monastery gardens, which revealed that units of heredity known as "genes" and gene variants called "alleles" govern hereditary traits.[1] Years later, scientists discovered deoxyribonucleic acid ("DNA") and cracked the genetic code.[2] Over the decades, researchers successfully isolated genes that predispose us to certain diseases and mapped the diseases to genetic markers by using restriction enzymes to cleave DNA at sequence-specific sites.[3] In October 1990, the United States Department of Energy ("DOE") and the National Institute of Health ("NIH") launched the Human Genome Project—an effort to map and sequence the human genome.[4] "As our knowledge and understanding of genetic information evolves, so, too, must the laws governing its use."[5]

1. *See infra* notes 22–23 and accompanying text.
2. *See infra* notes 37–38 and accompanying text.
3. *See infra* notes 43–48 and accompanying text.
4. *See infra* notes 50–51 and accompanying text.
5. Lawrence Z. Lorber & Abigail L. Perdue, *A Legal Evolution: The Influence of Genetics in the American Workplace*, HR ADVISOR, Sept./Oct. 2008, at 12, 12.

In response to growing concern over genetic discrimination in employment and insurance, former President George W. Bush signed the Genetic Information Nondiscrimination Act ("GINA") into law on May 21, 2008.[6] On November 9, 2010, the Equal Employment Opportunity Commission ("EEOC") published GINA's final regulations, which took effect on January 10, 2011.[7]

GINA prohibits covered entities, including health insurers and some employers, from discriminating on the basis of genetic information.[8] GINA aims to dispel fears that undergoing genetic testing or participating in genetic research will endanger privacy or interfere with one's ability to obtain insurance or employment.[9] GINA is yet another reminder of the increasing influence that genetics has and will continue to have on the American workplace.[10]

If you think that genetic discrimination could never personally affect you, think again.[11] According to one source, scientists have identified at least 15,500 genetic disorders that impact approximately thirteen million Americans.[12] By some estimates, "every human being carries genetic markers for anywhere from five to fifty serious disorders."[13] Because of these genetic predispositions, each of us could be a victim of genetic discrimination.[14]

Anecdotal evidence suggests that genetic discrimination has already adversely impacted some Americans. In 2000, two insurance companies purportedly rejected one woman's applications for life insurance after she revealed that her biological relatives suffered from a genetic disorder known as Huntington's disease, even though she had never been tested for the gene that causes it.[15] In another example, a woman decided to undergo a prophylactic ovariectomy and mastectomy after discovering that she carried

---

6. *Id.*

7. Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-233, 122 Stat. 881 (2008) (codified as amended at 42 U.S.C. §§ 2000ff–2000ff-12 (Supp. III 2009)).

8. *See infra* notes 96–107 and accompanying text.

9. Lorber & Perdue, *supra* note 5, at 12.

10. *See, e.g.*, Louise McIntosh Slaughter, *Genetic Testing and Discrimination: How Private Is Your Information?*, 17 STAN. L. & POL'Y REV. 67, 72–74 (2006) (discussing laws addressing discrimination based on genetic information).

11. *Id.* at 69; *see also* Ed Timms, *'Genetic Discrimination' Condemned; Bush, Lawmakers Favor Protection in Employment, Health Insurance*, DALL. MORNING NEWS, June 24, 2001, at 4A ("'Every blessed one of us has bad genes' . . . . 'If genetic discrimination is allowed each one of [us] is potentially uninsurable or unemployable.'").

12. Slaughter, *supra* note 10, at 67.

13. *Id.* For a general discussion of the broader implications of genetic discrimination in other areas of law, such as insurance, personal injury litigation, education, and family law, *see* Lori Andrews, *Body Science*, A.B.A. J., Apr. 1997, at 44, 45.

14. Slaughter, *supra* note 10, at 69.

15. *Id.* at 70.

the BRCA-1 gene, which indicates a predisposition to certain cancers.[16] After receiving the bill for the genetic testing, the woman's employer-based health insurer allegedly terminated her coverage.[17] She then lost her job, despite having received a series of positive performance evaluations.[18] Another individual claimed that he was denied employment after his pre-employment physical exam revealed his sex chromosome disorder.[19] One woman admitted, "I'm afraid if [my predisposition for Huntington's disease] gets into my personnel file, it'll never go away . . . I would be terrified of my job security . . . . I worry about it."[20]

President George W. Bush provided the following justification for GINA during legislative debate regarding the Act:

> Genetic discrimination is unfair to workers and their families. It is unjustified—among other reasons, because it involves little more than medical speculation. A genetic predisposition toward cancer or heart disease does not mean the condition will develop. To deny employment or insurance to a healthy person based only on a predisposition violates our country's belief in equal treatment and individual merit.[21]

Something had to be done, but was Title II of GINA the best solution?

This Article aims to answer that question by analyzing the new law and exploring the controversy surrounding its enactment. Section II provides a basic understanding of the genetic science at the heart of the GINA controversy and illustrates that, due to the exceptional nature of genetic information, genetic discrimination in employment is best addressed by genetic-specific legislation like GINA. Section III illustrates the existence of genetic discrimination in employment and demonstrates how the similarities between genetic discrimination, racism, and sexism support GINA's enactment. Section IV discusses how GINA strikes a balance

---

16. Tara L. Rachinsky, Comment, *Genetic Testing: Toward a Comprehensive Policy to Prevent Genetic Discrimination in the Workplace*, 2 U. PA. J. LAB. & EMP. L. 575, 575–76 (2000).

17. *Id.* at 576.

18. *Id.*

19. Julie Shoop, *Law and the Laboratory: Genetic Research Prompts Concerns About Bias*, TRIAL, Mar. 1997, at 12, 12.

20. *Genetic Testing Raises Fears of Workplace Bias*, DALL. MORNING NEWS, Apr. 26, 1998, at 8H.

21. *The Potential for Discrimination in Health Insurance Based on Predictive Genetic Tests: Hearing Before the Subcomm. on Commerce, Trade, and Consumer Prot. of the H. Comm. on Energy and Commerce*, 107th Cong. 12 (2001) (statement of Rep. Constance A. Morella), *available at* http://www.access.gpo.gov/congress/house; *see White House Seeks a Ban on 'Unfair' Genetic Bias*, WASH. POST, June 24, 2001, at A8; *see also* Sheryl Gay Stolberg, *Panel Breaks Logjam for Bill on Employees' Genetic Histories*, N.Y. TIMES, May 22, 2003, at A21 (discussing approval of genetic discrimination bill after six years in a Senate committee).

between the competing interests regarding the use of genetic information and highlights countervailing employee concerns arising from the use of genetic testing in employment, such as worker autonomy and employees' fears of genetic stigmatization, medicalization, and discrimination. Finally, Section V suggests points to consider in future amendments to GINA and the promulgation or modification of GINA regulations.

II. GENETICS 101

A. The Evolution of Genetics

In the 1860s, Austrian monk Gregor Mendel observed that the flower color of pea plants in his monastery gardens exhibited predictable patterns of inheritance.[22] His experiments proved that functional and physical units of heredity—genes and alleles—govern hereditary traits.[23] Unfortunately, because Darwin's theory of evolution distracted the scientific community, Mendel's work was all but forgotten until three scientists—Hugo DeVries, Erich Von Tschermak, and Carl Correns—duplicated Mendel's experiments, renewing interest in his work.[24]

B. Decoding the Genetic Code

A full grasp of the problems inherent in policing genetic discrimination in employment requires a basic understanding of genetics, including its vocabulary. "Genotype" refers to an individual's genes, while "phenotype" is the physical manifestation of the genotype.[25] In a Mendelian pattern of inheritance, such as pea flower color, dominant traits are expressed even if only one copy of the dominant allele exists.[26] Recessive alleles must exist in two copies to be expressed.[27] "Homozygous" refers to two identical copies

22. Sue Goetinck, *DNA Mapping Could Revolutionize Medicine; Genetic Data May Spur Cures, Discrimination*, DALL. MORNING NEWS, July 19, 1998, at 1A.

23. *See* LARRY GONICK & MARK WHEELIS, THE CARTOON GUIDE TO GENETICS 37–50, 42 (Harper Perennial 1991) (1983); Heather Rae Watterson, *Genetic Discrimination in the Workplace and the Need for Federal Legislation*, 4 DEPAUL J. HEALTH CARE L. 423, 426 (2001). To illustrate this, imagine that gene *B* is the gene that codes for blue eyes. Its first form, Allele One, causes aquamarine blue eyes, while Allele Two causes sky blue eyes.

24. GONICK & WHEELIS, *supra* note 23, at 55; JACK J. PASTERNAK, AN INTRODUCTION TO HUMAN MOLECULAR GENETICS: MECHANISMS OF INHERITED DISEASE 7 (Fitzgerald Science Press 1999) (1980).

25. GONICK & WHEELIS, *supra* note 23, at 49.

26. *See id.* at 49–50, 54.

27. 2 ENCYCLOPEDIA OF BIOETHICS 971 (Stephen G. Post ed., 3d ed. 2004); *see, e.g.*, GONICK & WHEELIS, *supra* note 23, at 53–54. The following are examples of the relationship between dominant and recessive genes: brown eyes are dominant over blue eyes; color vision is dominant over color blindness; hairy heads are dominant over bald heads; and having extra fingers is dominant over having only five. *Id.* at 54. Many rare genetic disorders

of the same allele (*i.e.*, *AA* or *aa*), and "heterozygous" refers to a pair of non-identical alleles (*i.e.*, *Aa*).[28] "Autosomal" indicates that a gene is not located on a sex chromosome.

If you are wondering whom to blame for your curly hair or bald spot, just remember that each person inherits one copy of each gene from each parent. Different alleles are sorted out randomly via independent assortment to eggs and sperm during meiosis,[29] a process in which each sperm and each egg receive one copy of each gene.[30] To our current understanding, all allele combinations are equally likely.[31] Meiosis results in four sex cells (*i.e.*, sperm or egg), each with twenty-three chromosomes.[32] During fertilization, the nuclei of the sperm and egg merge, creating a zygote, or fertilized egg, that contains a full set of forty-six chromosomes—twenty-three from the father and twenty-three from the mother.[33]

After discovering genes and their patterns of inheritance, geneticists turned to the question of how genes work. The answer is simpler than one might expect. Each gene is responsible for directing the manufacture of one specific enzyme.[34] To do this, the gene relies on a special genetic code that was discovered in 1961 by scientist Marshall Nirenberg.[35] He revealed that the genetic code is the same in all life forms from plants to people.[36]

In 1944, Oswald Avery discovered that DNA inside each gene contains four bases—adenine ("A"), guanine ("G"), cytosine ("C"), and thymine ("T")—that exist in series of base triplets called "codons."[37] Codons act like words to code for specific amino acids.[38] For example, *ACA* codes for threonine; *AGU* codes for serine; *UUU* and *UUC* code for phenylalanine;

---

like sickle cell anemia, hemophilia, Tay-Sachs, thalassemia (an inability to make hemoglobin causing its victim to suffer a severe lack of oxygen), and dwarfism manifest only if the person has a homozygous recessive genotype (i.e., *aa*). *Id.*

28. GONICK & WHEELIS, *supra* note 23, at 49.

29. *Id.* at 61–64 (illustrating the process of meiosis). Sperm and egg are each gametes, or germ cells, that contain twenty-three chromosomes, which is half of the total chromosomes of a human. *Id.* at 61. In 1902, American scientist William Sutton discovered "that each chromosome from the sperm can be matched with a virtually identical one from the egg," meaning that the zygote contains twenty-three homologous chromosomal pairs. *Id.* The only exception is the pair of sex chromosomes, which is XX for women and XY for men. *Id.* The homolog donated to the zygote is completely random due to a phenomenon called "independent assortment." *Id.* at 64.

30. *Id.* at 54.

31. *Id.*

32. *Id.* at 64

33. *Id.* at 62.

34. PASTERNAK, *supra* note 24, at 14.

35. GONICK & WHEELIS, *supra* note 23, at 134.

36. *Id.* at 145.

37. *Id.* at 118–20, 134.

38. *Id.* at 134–35.

and *AUG* codes for methionine.[39] Three codons known as "stop codons" signal the end of a protein chain.[40] Once produced, the amino acids join to form a chain called a "peptide," and peptides join to form a protein.[41] Though most proteins are enzymes, proteins also compose human fingernails, hair, and even bird feathers.[42]

Scientists soon began to wonder how to map gene locations on a chromosome. They discovered that genes cross over, switching spots from Homolog *A* to Homolog *B* in Pair One.[43] Cross-over frequency increases with distance, allowing scientists to map gene locations on chromosomes.[44] Genes that usually cross together are said to be "linked," meaning that they are located very close together on the chromosome.[45]

Scientists also discovered restriction fragment length polymorphisms ("RFLPs"), which are inherited DNA sequences that can be used as genetic markers to map genetic diseases and to document their inheritance through families.[46] In 1983, Huntington's disease[47] became the first disease to be mapped to a genetic marker through the use of restriction enzymes at sequence-specific sites.[48] After researchers successfully isolated the gene for cystic fibrosis in 1989, the Human Genome Project began in 1990.[49]

C. The Human Genome Project

In October 1990, the DOE and NIH launched the Human Genome Project ("HGP")—"a coordinated, international research effort . . . to map and sequence the estimated 50,000 to 100,000 human genes."[50] In April

---

39. *See id.* at 135 (chart of amino acids produced from certain codon combinations).

40. *Id.* at 135.

41. *Id.* at 111.

42. *Id.* at 111; *see also id.* at 138–45 (discussing how proteins are made).

43. *Id.* at 77.

44. *Id.*

45. *See id.*

46. 2 ENCYCLOPEDIA OF BIOETHICS, *supra* note 27, at 1021.

47. Huntington's disease is "a late-onset autosomal dominant neuropsychiatric disorder" that leads to diminished mental cognition, severe mood swings, and a variety of other symptoms. *Id.*

48. *Id.* Huntington's disease typically appears in an individual's forties and has no known cure. *Id.*

49. Goetinck, *supra* note 22, at 1A.

50. Paul Steven Miller, *Genetic Discrimination in the Workplace*, 26 J.L. MED. & ETHICS 189, 194 n.1 (1998); *see also* Jennifer R. Taylor, *Mixing the Gene Pool and the Labor Pool: Protecting Workers from Genetic Discrimination in Employment*, 20 TEMP. ENVTL. L. & TECH. J. 51, 53 (2001) ("The Human Genome Project (HGP) is an international 13-year effort that began in October 1990 in an effort to determine all the estimated 30,000–35,000 human genes.").

2003, the HGP announced a completed, high-quality DNA reference sequence.[51]

The HGP has evoked mixed public reaction. Supporters claim it will improve biomedical research and lead to medical advances, such as prophylactic treatments for a wide variety of genetically-linked conditions.[52] Yet, a poll taken in June 2000 revealed that 46% of participants thought that the HGP would yield "hazardous results" and over 40% considered it to be "morally wrong."[53]

D. From Pea Plants to Punnett Squares: Understanding Inheritance

Any discussion of genetic discrimination in employment is incomplete without a comprehensive explanation of the complicated nature of inheritance. In fact, it is often a misplaced faith in "genetic determinism"[54] that causes employers and the public at large to engage in largely irrational genetic discrimination and stigmatization.

To begin, consider the following illustration of a Mendelian pattern of inheritance.[55] Imagine that one autosomal gene determines hair color. *A* is an allele that causes red hair, while *a* causes blond hair. A person with an *AA* genotype possesses a homozygous dominant genotype and expresses a dominant phenotype, namely red hair. If two homozygous dominant redheads mate, they will produce offspring with an *AA* genotype and red hair. *Aa* is an example of a heterozygous genotype, meaning that the individual contains one dominant allele and one recessive allele. An individual with an *Aa* genotype will still have red hair because the dominant allele's expression masks the recessive allele's expression. Therefore, a person possessing an *AA* genotype will appear the exact same way as a person with an *Aa* genotype, even though the latter carries a recessive allele. If one *Aa* mates with another *Aa*, there is a 25% chance that the couple will produce an *AA* redhead, a 50% chance that the couple will produce an *Aa* heterozygous redhead, and a 25% chance that the couple will produce an *aa* homozygous recessive blond. If a blond male mates with a blonde female, they have a 100% chance of producing blond offspring; however, if a blond mates with an *AA* redhead, all of the children will be heterozygous redheads, each of whom will carry a recessive *a* allele.

---

51. 2 ENCYCLOPEDIA OF BIOETHICS, *supra* note 27, at 1021.

52. Miller, *supra* note 50, at 194 n.1.

53. Tobi T. Bromfield, *Your DNA Is Your Resume: How Inadequate Protection of Genetic Information Perpetuates Employment Discrimination*, 7 WASH. & LEE RACE & ETHNIC ANC. L.J. 117, 123 (2001).

54. *See infra* notes 94–95 and accompanying text.

55. To better understand the given illustration, *see* the full discussion of genetics terminology and Mendelian inheritance in notes 25–33 and accompanying text in this Article.

Not all patterns of genetic inheritance are this simple. Instead, genetic inheritance is often complicated by various phenomena, including but certainly not limited to mutation, codominance, sex linkage, heritability, and ecogenetics.

E. Mutation and Codominance

Genetic mutation is a very rare phenomenon that is usually triggered by environmental influences. Each one of us potentially carries a number of mutant genes, and while some mutant genes can cause defects, most mutations are not expressed.[56] Codominance occurs when a heterozygote individual expresses both phenotypes of each different allele.[57] Blood groups are illustrative: while a homozygous *AA* person has Type A blood and a homozygous *BB* individual has Type B blood, a heterozygous *AB* person has Type AB blood.[58] The *AB* individual produces the *A* phenotype (*i.e.*, Type A antibodies) as well as the *B* phenotype (*i.e.*, Type B antibodies).[59] Neither allele masks the other.[60]

F. Sex Linkage

If you have ever wondered why most bald people are men, you can attribute that to a phenomenon called "sex linkage."[61] Likewise, color blindness and hemophilia[62] typically afflict men.[63] Such sex-linked traits occur because each human male inherits one X chromosome from his mother and one Y chromosome from his father.[64] By contrast, each human female contains a homologous pair of sex chromosomes: an X chromosome from her mother and an X chromosome from her father.[65] Men and women share the other twenty-two pairs of autosomal chromosomes.[66] Because a man only receives one X chromosome from his mother and none from his father, any recessive alleles on his X chromosome will be expressed, including the recessive alleles that cause sex-linked traits like male pattern

---

56. GONICK & WHEELIS, *supra* note 23, at 79–80.
57. *Id.* at 162–63.
58. *Id.*
59. *Id.*
60. *See id.* Another example is sickle cell anemia, which is codominant rather than recessive.
61. *Id.* at 91.
62. Hemophilia is a failure of the blood to clot. *Id.*
63. *Id.*
64. *Id.* at 92.
65. *Id.*
66. *Id.* at 62, 91.

baldness, color blindness, and hemophilia.[67] This occurs because a man lacks a dominant allele to mask the expression of his recessive alleles.

To illustrate sex linkage, consider this situation: Bald Jerry ($X^{bald}Y$) marries Normal Kate (XX).[68] Their daughters, Jane and Sue, will be carriers of the "bald" allele though neither will be bald.[69] If Carrier Jane ($X^{bald}X$) marries Normal Jon (XY), half of their daughters will be carriers of the "bald" gene and half of their sons will be bald.[70] A more famous example of sex linkage involves Great Britain's Queen Victoria. As the story goes, a recessive allele for hemophilia—aptly nicknamed "the royal disease"—spontaneously arose in her genes via mutation. Consequently, her son, three of her grandsons, and six of her great-grandsons were hemophiliacs.[71]

G. Heritability

"Heritability" refers to the percentage of genetic causation for a genetic trait as determined via twin concordance studies.[72] By relying on the "equal environments assumption,"[73] scientists have determined the heritability of various genetically-influenced conditions.[74] According to one study, the heritability of schizophrenia is 45% for identical twins, 17% for fraternal twins, and 9% for full siblings.[75] Despite that, 63% of schizophrenics have neither a first- nor second-degree relative with the disorder.[76] This concordance pattern reveals that for identical twins, more than half of the difference in the likelihood of developing schizophrenia is due to non-shared environmental factors—the most relevant of which appear to

---

67. *Id.* at 91–92.

68. *Id.* at 93.

69. *Id.*

70. *Id.*

71. *See id.* at 95.

72. 2 ENCYCLOPEDIA OF BIOETHICS, *supra* note 27, at 972. Twin studies are especially useful because monozygotic twins share 100% of the same genes because they were formed when one fertilized egg split to become two embryos. *See* PASTERNAK, *supra* note 24, at 12–13. Dizygotic twins share 50% of their genes but share the same intrauterine environment, which likely accounts for their higher heritability when compared with full siblings who also share 50% of their genes but do not share an intrauterine environment. *See id.*

73. 2 ENCYCLOPEDIA OF BIOETHICS, *supra* note 27, at 972. The "equal environment assumption" assumes that twins raised together have the same environment and those raised apart have different environments. The meaning of environment in quantitative genetics is extremely broad, denoting everything that is not genetic. *Id.* "The shared environment comprises all the nongenetic factors that cause family members to be similar, and the nonshared environment is what makes family members different." *Id.* at 973.

74. *Id.*

75. *Id.*

76. *Id.* at 973.

include, *inter alia*, problems in pregnancy, obstetric complications, and urban or winter birth.[77]

Schizophrenia is not the only disorder subject to strong genetic influences. To the contrary, one source indicates that among the eleven personality scales, 54% of the variation is attributable to genetic differences and 46% to environmental differences.[78] By some estimates, 70% of the difference in human intelligence is genetic and 30% environmental.[79] Alcoholism also runs in biological families, and 25% of men with alcoholic relatives become alcoholics.[80] For this reason, a child without alcoholic biological relatives has no increased risk of developing alcoholism even if raised by alcoholic adoptive parents, which indicates that genes play a larger role than environment.[81]

H. Ecogenetics

A monogenic disease is foretold by the presence of a single mutation in a single gene, but as the heritability of schizophrenia illustrates, many, if not most, genetic disorders and genetically influenced conditions are not monogenic.[82] Diseases more typically arise from the interaction between an organism's genome and its environment, including diet, climate, and intrauterine environment.[83] Different genotypes have different sensitivities to different environments, further complicating the predictive value of genetic testing.[84]

Ecogenetics is the study of this gene–environment interaction and is relevant to determining which persons are more likely to express a genotype's corresponding phenotype.[85] The Environmental Genome Project ("EGP") conducts epidemiological studies to investigate the role of genome–environment interactions in the development of common diseases like asthma, cancer, and heart disease, and to promote the use of that information in public health initiatives.[86]

One condition particularly illustrative of the genome-environment interaction is phenylketonuria ("PKU"), a congenital autosomal recessive

77. *Id.*
78. *Id.* at 973.
79. *Id.* at 972.
80. *Id.* at 973.
81. *Id.*
82. *Id.* at 956.
83. *Id.*
84. *Id.*
85. *Id.* at 967.
86. *Id.* at 968. The EGP is especially important in emphasizing that a person's genetic susceptibility to one type of environmental stimulus does not mean that the individual must be labeled as "hypersusceptible" to every similar stimulus. Moreover, the EGP will clarify that although a person may have a genetic risk factor, the cause of the disease will still be unclear; it may or may not have been due to a genetic predisposition.

condition that causes severe mental retardation, seizures, temper tantrums, and early death.[87] If detected at birth via a blood prick test on the infant's foot, the child is placed on a special phenylalanine-free diet that will prevent his PKU from being expressed.[88] The diet prevents or at least minimizes the nervous system damage that untreated PKU victims would otherwise endure.[89]

Genetic predispositions to cancer also demonstrate EGP's positive effects. According to one source, a woman with no predisposition to breast cancer has a 13% risk of developing it compared to the 56% risk that a genetically-predisposed woman has.[90] Absent a genetic predisposition for developing ovarian cancer, a woman has a 1.6% risk of developing it, but a woman with any one of the three mutations linked to ovarian cancer has ten times that risk.[91] Males who are genetically predisposed to prostate cancer run a 16% risk of developing it compared to a 3.8% risk for males who are not predisposed.[92] For the general public, the risk of developing colon cancer is 5%, but for Ashkenazi Jews who carry a mutation that predisposes them to colon cancer, the risk of development is 18 to 30%.[93]

I. Dominance or Destiny

"Genetic determinism" refers to the view that an individual's fate lies in his or her genes—the idea that an individual who possesses a gene for cancer *will* develop the cancer.[94] Unfortunately, many people, including employers, erroneously subscribe to genetic determinism. In one instance, a father allegedly insisted that his ten-year-old daughter undergo a full ovariectomy and mastectomy because she carried a gene that predisposed her to breast cancer.[95] Like so many others, her father failed to understand that carrying the gene was not a death sentence; it simply meant that his daughter had a heightened susceptibility to cancer. Even without the operations, she may never have developed cancer and could have led a "normal" life, possibly even bearing children. Her father's well-intentioned but perhaps misguided decision may have robbed her of that choice because

87. *Id.* at 972.

88. *Id.* Phenylalanine is the amino acid that is the environmental trigger, which causes the genetic condition to express itself.

89. *Id.*

90. Sheryl Gay Stolberg, *Concern among Jews Is Heightened as Scientists Deepen Gene Studies,* N.Y. TIMES, Apr. 22, 1998, at A24.

91. *Id.*

92. *Id.*

93. *Id.*

94. *See generally* Jane Maienschen, *Cloning and Stem Cell Debates in the Context of Genetic Determinism*, 9 YALE J. HEALTH POL'Y L. & ETHICS 565, 572–76 (defining and discussing the concept of genetic determinism).

95. Leon R. Kass, *The Age of Genetic Technology Arrives*, AM. SPECTATOR, Nov./Dec. 2002, at 40, 42.

he failed to understand the subtle difference between a genetic predisposition and an expressed genetic disorder. Likewise, employers who engage in genetic discrimination unjustly deny prospective and present employees opportunities to reach their full career potential.

III. EXPLORING TITLE II

Misinformation about genetics and the misguided decision-making and discrimination to which it often leads prompted GINA's enactment. GINA's drafters describe it as civil rights legislation designed to eliminate Americans' fears of genetic discrimination in employment and insurance.[96] Title I of GINA, which relates to health insurance, took effect in May 2009.[97] It expands existing anti-discrimination provisions in the Health Insurance Portability and Accountability Act ("HIPAA") by amending the Employee Retirement Income Security Act of 1974 ("ERISA"), the Public Health Service Act, the Internal Revenue Code of 1986, and Title XVIII of the Social Security Act.[98]

Title II of GINA, which relates to employment, became effective in November 2009, and its accompanying regulations, which were adopted in November 2010, took effect on January 10, 2011.[99] Title II, which is the sole focus of this Article, prohibits covered entities[100]—certain employment agencies, employers, joint labor–management committees, and labor organizations—from discriminating against a covered applicant or current or former employee[101] on the basis of genetic information.[102] Covered

---

96. Lauren J. Sismondo, Note, *GINA, What Could You Do for Me One Day?: The Potential of the Genetic Information Nondiscrimination Act To Protect the American Public*, 21 WASH. U. J.L. & POL'Y 459, 472–73 (2006).

97. Lorber & Perdue, *supra* note 5, at 12.

98. *Id.* at 12–13.

99. *Id.* at 13; *see supra* note 7 and accompanying text.

100. "[A]n employer does not include an Indian tribe, or a bona fide private club (other than a labor organization) that is exempt from taxation under section 501(c) of the Internal Revenue Code of 1986." Genetic Information Nondiscrimination Act of 2008, 29 C.F.R. § 1635.2(d) (2011).

101. GINA defines employee in reference to other federal laws:

> The term 'employee' means (i) an employee (including an applicant), as defined in section 2000e(f) of this title; (ii) a State employee (including an applicant) described in section 2000e-16c(a) of this title; (iii) a covered employee (including an applicant), as defined in section 1301 of Title 2; (iv) a covered employee (including an applicant), as defined in section 411(c) of Title 3; or (v) an employee or applicant to which section 2000e-16(a) applies.

42 U.S.C. § 2000ff(2) (Supp. III 2009); *see also* 29 C.F.R. § 1635.2(c) (noting that GINA also applies to former employees).

102. 42 U.S.C. § 2000ff-1.

entities may not (i) base employment, membership, or training decisions on genetic information;[103] (ii) limit, segregate, or classify people on the basis of genetic information in a way that deprives or tends to deprive them of employment opportunities or adversely impacts their employment status;[104] (iii) retaliate against any individuals for exercising their rights under GINA;[105] or (iv) request, require, purchase, or disclose genetic information, except in limited circumstances.[106] With regard to retaliation, GINA specifically prohibits a covered entity from discriminating against an individual because he or she opposed any conduct or practice that violates the statute or made a "charge, testified, assisted, or participated in any manner in [a GINA] investigation, proceeding, or hearing."[107]

GINA defines "genetic information" as information about genetic tests of an individual and that individual's family members as well as information about any family member's "manifested"[108] disease or disorder.[109] Genetic information also encompasses individuals' and their family members' requests for or receipt of genetic services,[110] including testing, counseling, education, and participation in clinical research that involves such services.[111] This definition expressly excludes information about sex and age but includes genetic information about a fetus carried by a covered individual, by an individual's pregnant family member, or an

---

103. Id. § 2000ff-1(a)(1); Lorber & Perdue, *supra* note 5, at 13.

104. 42 U.S.C. § 2000ff-1(a)(2); Lorber & Perdue, *supra* note 5, at 13.

105. 42 U.S.C. § 2000ff-6(f) ("No person shall discriminate against any individual because such individual has opposed any act or practice made unlawful by this chapter or because such individual made a charge, testified, assisted, or participated in any manner in an investigation, proceeding, or hearing under this chapter."); Lorber & Perdue, *supra* note 5, at 13.

106. 42 U.S.C. § 2000ff-1(b); Lorber & Perdue, *supra* note 5, at 13.

107. Genetic Information Nondiscrimination Act of 2008, 29 C.F.R. § 1635.7 (2011).

108. GINA regulations provide a definition for manifested conditions:

> *Manifestation* or *manifested* means . . . that an individual has been or could reasonably be diagnosed with the disease, disorder, or pathological condition by a health care professional with appropriate training and expertise in the field of medicine involved. For purposes of this part, a disease, disorder, or pathological condition is not manifested if the diagnosis is based principally on genetic information.

Id. § 1635.3(g).

109. 42 U.S.C. § 2000ff(4); Regulations Under the Genetic Information Nondiscrimination Act of 2008, 74 Fed. Reg. 9056, 9060 (proposed Mar. 2, 2009) (to be codified at 29 C.F.R. pt. 1635) ("[E]ven when a genetic variant is 100 percent predictive for development of disease, the presence of the variant does not by itself equal diagnosis of the disease.").

110. 42 U.S.C. § 2000ff(4)(B).

111. *Id.* § 2000ff(4)(B), (6).

embryo lawfully held by a covered individual or family member who is undergoing assisted reproduction.[112]

Notably, GINA does not cover a person's manifested condition, even if it is genetically linked.[113] Such coverage remains the exclusive purview of the Americans with Disabilities Act ("ADA") as amended.[114] GINA's accompanying regulations define "manifested" as "a disease, disorder, or pathological condition, that an individual has been or could reasonably be diagnosed with . . . by a health care professional with appropriate training and expertise in the field of medicine involved."[115]

The following example illustrates how GINA applies to an employee's manifested genetic disorder as compared to the manifested genetic disorder of the employee's family member: the results of a genetic test indicating that employee Jane Doe carries an altered version of the BRCA-1 gene and the fact that Jane's sister has breast cancer constitute protected genetic information under GINA, but the results of a biopsy indicating that Jane actually has breast cancer are not protected.[116] Consequently, GINA permits Jane's employer to use, acquire, or disclose non-genetic medical information about her manifested condition (even though her condition may have a genetic basis).[117] In other words, even if Jane had a genetic predisposition to Carpal Tunnel Syndrome, her employer would not violate GINA by reporting her manifested Carpal Tunnel condition to its workers' compensation carrier, because this is disclosing phenotypic—not genotypic—information.[118]

GINA broadly defines "family member" to include dependents[119] as a result of marriage, birth, adoption, or placement for adoption as well as any other first-, second-, third-, or fourth-degree relatives.[120] In other words, the term "family member" encompasses an individual's children, siblings, parents, great-great grandparents, first cousins once removed, and everyone in between, including half-siblings and adopted relatives.[121]

---

112. *Id.* §§ 2000ff(4)(C), 2000ff-8(b).

113. *Id.* § 2000ff-9 (allowing employer's "use, acquisition, or disclosure of a medical information that is not genetic information about a manifested disease, disorder, or pathological condition of an employee or member, including a manifested disease, disorder, or pathological that *has or may have a genetic basis*") (emphasis added).

114. *See* Genetic Information Nondiscrimination Act of 2008, 29 C.F.R. § 1635.12(a)(2) (2011).

115. *See supra* note 108 and accompanying text.

116. *See supra* notes 113–115 and accompanying text.

117. *See supra* notes 113–115 and accompanying text.

118. *See supra* notes 113–115 and accompanying text.

119. GINA derives its definition of dependent from Section 701(f)(2) of ERISA. Regulations under the Genetic Information Nondiscrimination Act of 2008, 74 Fed. Reg. 9056, 9058 (proposed Mar. 2, 2009) (to be codified at 29 C.F.R. pt. 1635).

120. 42 U.S.C. § 2000ff(3) (Supp. III 2009); Genetic Information Nondiscrimination Act of 2008, 29 C.F.R. § 1635.3(a) (2011).

121. Regulations under the Genetic Information Nondiscrimination Act of 2008, 74

To the extent "family member" includes spouses and adopted relatives, the definition of family member is overbroad because an adopted person would not likely share the same DNA as his adopted family and spouses presumably are not genetically similar. As such, there is no reason why the term "family member" under GINA should extend to adopted relatives and spouses. Even if the adoptee is also a biological relative—such as when an aunt legally adopts her niece or nephew after the child's parents pass away—the inclusion is redundant because GINA would still cover the aunt–niece biological relationship regardless of the adoption.

GINA defines "genetic test" as an "analysis of human DNA, RNA, chromosomes, proteins, or metabolites that detects genotypes, mutations, or chromosomal changes."[122] GINA's accompanying regulations clarify that the term "genetic test" includes, *inter alia*, tests to determine a genetic predisposition for a disorder or genetic variant; amniocentesis to test a fetus during pregnancy; some types of newborn screening; pre-implantation genetic diagnosis performed on embryos created via *in vitro* fertilization; pharmacogenetic tests that detect genotypes, mutations, or chromosomal changes; DNA testing that detects genetic markers associated with information about ancestry; and DNA testing that aims to reveal family relationships (*i.e.*, paternity tests).[123]

However, an analysis that does not detect genotypes, mutations, or chromosomal changes is not a genetic test under GINA.[124] For example, tests meant solely to detect the presence of a virus do not constitute a genetic test.[125] Routine liver function tests, blood counts, or cholesterol tests along with tests for infectious or communicable diseases that could be transmitted through food handling are also excluded from GINA's protection.[126] While a test for the presence of alcohol or drugs is not a genetic test, a test to determine whether an individual possesses a genetic predisposition for alcoholism or drug use does constitute a genetic test.[127]

Prior to GINA's enactment, employers primarily used two types of genetic testing to screen employees: genetic screening and genetic monitoring.[128] Genetic screening can determine an applicant or employee's genetic predisposition and occurs via biochemical genetic screening or

---

Fed. Reg. at 9058–59; *see* 29 C.F.R. § 1635.3(a).

122. 42 U.S.C. § 2000ff(7)(A).

123. 29 C.F.R. § 1635.3(f)(2).

124. *Id.* § 1635.3(f)(3).

125. *Id.* § 1635.3(f)(3)(ii), (f)(4)(i).

126. *Id.* § 1635.3(f)(3)(iii)–(iv).

127. *Id.* § 1635.3(f)(4)(ii).

128. Jared A. Feldman & Richard J. Katz, Note, *Genetic Testing & Discrimination in Employment: Recommending a Uniform Statutory Approach*, 19 HOFSTRA LAB. & EMP. L.J. 389, 395 (2002).

direct DNA screening.[129] DNA screening directly examines a donor's DNA sample.[130]

A market for genetic tests has already developed. In June 2002, Myriad Genetics began marketing genetic tests for certain cancers to the general public, even though the tests were only appropriate for use by a very small portion of the population.[131] In another case, a test for the APOE e4 allele, which is associated with late-onset Alzheimer's disease, was sold directly to physicians before research was completed to determine how to interpret the connection between the existence of the allele and development of the condition.[132] In 2003, a San Francisco firm began selling genetic tests for cancer, cystic fibrosis, hemachromatosis,[133] and other diseases to the public.[134] Another company marketed a test for variations in a serotonin transporter gene linked to chronic depression, and a laboratory allegedly sold tests for genes linked to macular degeneration and glaucoma for $99.95.[135]

However, genetic testing is not foolproof. Rather, it is complicated by heterogeneity and penetrance. "Heterogeneity" refers to "the fact that the same genetic disease may result from the presence of any of several different variants of the same gene . . . or of different genes."[136] For instance, various forms of cystic fibrosis ("CF"), which differ in severity, can be caused by any one of 900 potential mutations of the CF gene.[137] "Penetrance" is the probability that a genetic disease will manifest symptoms when the disease-related genotype exists.[138]

To further complicate matters, screening tests also have high rates of false positives,[139] and some laboratories have failed quality and proficiency assessments. Other mix-ups, including human error, also affect the reliability of results obtained. A related problem is that recipients of testing may misinterpret the meaning of a positive result, perhaps mistaking increased susceptibility to a disease with actually having the disorder.[140]

---

129. *Id.*

130. *Id.*

131. 2 ENCYCLOPEDIA OF BIOETHICS, *supra* note 27, at 1022.

132. *Id.*

133. *See* Nancy Shute, *Unraveling Your DNA's Secrets*, U.S. NEWS & WORLD REP., Jan. 8, 2007 at 50, 52 ("[G]enetic testing has become absurdly simple: Buy a test online, and within a few days a kit arrives in the mail. Rub the small brush on the inside of your cheek for 30 seconds, pop it back in the prepaid envelope, mail it back, and *voilà*! In a short time, you'll receive the truth about your genes.").

134. *See* DNA DIRECT, http://www.dnadirect.com/web/ (last visited January 9, 2012).

135. Shute, *supra* note 133 at 53–54.

136. 2 ENCYCLOPEDIA OF BIOETHICS, *supra* note 27, at 1021.

137. *Id.*

138. *Id.*

139. *Id.* at 998.

140. *See* Shute, *supra* note 133 at 57–58 ("Genetics is notoriously confusing and most people need help interpreting the test results. Most primary care doctors never studied

Alternately, some tests check for only one gene causally related to a genetic disorder, even though many other genes may also cause that disorder.[141] To the extent a donor has no control over what happens to his tissue sample and there is no doctor–patient confidentiality privilege applicable to genetic lab technicians, an opportunistic employer could easily pay a loquacious lab technician to provide him with the results of his employee's genetic test. Without GINA, the employee may have had no legal recourse for discriminatory employment decisions based on the genetic information obtained. Furthermore, direct-to-consumer genetic tests are not FDA-regulated because they are considered "services,"[142] and their results usually do not become part of a person's medical record.[143]

Some speculate that genetic tests will eventually become available to predict predispositions for a vast array of conditions, including diabetes, asthma, obesity, dyslexia, and schizophrenia.[144] Clearly then, in the absence of GINA, employers may have been able to purchase genetic tests for a wide variety of conditions and institute in-office genetic testing for almost anything.

Turning from statutory definitions to practical implications, GINA prohibits covered entities from requesting, requiring, or buying genetic information.[145] GINA defines "request" to include such acts as performing an Internet search on a person that is likely to reveal genetic information, actively eavesdropping on third-party conversations, searching through someone's personal effects to uncover genetic information, and requesting information regarding someone's health status in a manner that will probably disclose genetic information.[146] Thus, GINA could be interpreted broadly to bar a request even where genetic information is ultimately not obtained.

Limited exceptions to this rule do exist. For example, an inadvertent request for the family medical history of a covered individual or a family member or acquisition of family history through the purchase of commercially and publically available documents[147] does not violate GINA.[148] Furthermore, if a covered entity accidentally receives genetic

---

genetics in medical school, and there are only about 800 U.S. physicians board-certified in genetics.").

141. *See id.* at 57.

142. *Id.* at 54.

143. *See id.* at 57.

144. 2 ENCYCLOPEDIA OF BIOETHICS, *supra* note 27, at 1021.

145. 42 U.S.C. § 2000ff-1(b) (Supp. III 2009).

146. 29 C.F.R. § 1635.8 (2011).

147. *See id.* § 1635.8(b)(4). The statute identifies newspapers, magazines, periodicals, books, and electronic media (Internet, television, and movies) as commercially and publicly available. However, the statute excludes medical databases, court records, limited-access websites, and online discussion groups that focus on issues like genetic testing or genetic discrimination. *Id.*

148. 42 U.S.C. § 2000ff-1(b)(1), (b)(4); *see* Regulations under the Genetic Information

information as a result of a lawful request for medical information, no GINA violation has occurred unless the covered entity failed to direct the healthcare provider (either in writing or verbally) not to disclose genetic information.[149] In fact, an acquisition of genetic information as a result of the request is deemed inadvertent so long as the covered entity used the following language in its request:

> The Genetic Information Nondiscrimination Act of 2008 (GINA) prohibits employers and other entities covered by GINA Title II from requesting or requiring genetic information of an individual or family member of the individual, except as specifically allowed by this law. To comply with this law, we are asking that you not provide any genetic information when responding to this request for medical information. "Genetic information" as defined by GINA, includes an individual's family medical history, the results of an individual's or family member's genetic tests, the fact that an individual or an individual's family member sought or received genetic services, and genetic information of a fetus carried by an individual or an individual's family member or an embryo lawfully held by an individual or family member receiving assistive reproductive services.[150]

Even if a covered entity fails to give notice or to use this specific language, the employer can still show that the acquisition of genetic information was inadvertent if its request for medical information was not likely to result in the acquisition of genetic information.[151] The regulations accompanying GINA also emphasize that employers may still request medical documentation to support a request for a reasonable accommodation where the disability or need for the accommodation is not obvious and the documentation requested is not more than is necessary to establish the legitimacy of the accommodation request.[152]

The inadvertence exception also applies where a manager, supervisor, union representative, or employment agency representative acquires genetic information "by overhearing a conversation between the individual and others" or "by receiving it from the individual or third-parties during a casual conversation, including in response to an ordinary expression of concern that is the subject of the conversation."[153] However, the exception

---

Nondiscrimination Act of 2008, 74 Fed. Reg. 9056, 9061 (proposed Mar. 2, 2009) (to be codified at 29 C.F.R. pt. 1635) ("Congress did not want casual conversation among co-workers regarding health to trigger federal litigation whenever someone mentioned something that might constitute protected family medical history.").

149. 29 C.F.R. § 1635.8(b)(1)(A).

150. *Id.* § 1635.8(b)(1)(i)(B).

151. *Id.* § 1635.8(b)(1)(i)(C).

152. *Id.* § 1635.8(b)(1)(D).

153. *Id.* § 1635.8(b)(1)(ii)(A)–(B).

does not apply where a manager follows up the inadvertent disclosure with more probing questions that are likely to reveal genetic information.[154]

By way of illustration, imagine the following two scenarios. In Scenario One, Manager Smith asks Employee Jane, "How are you doing today?" Jane responds, "I'm not doing very well today. We just got the amniocentesis results back, and there is something wrong with the baby." Manager Smith responds, "I'm sorry to hear that, Jane. What's wrong?" Jane responds, "He has Down Syndrome." Manager Smith responds, "Does that run in your family?" Jane responds, "Yes. I have two cousins and one nephew who also have it. Now we are afraid all of our children might have it if we decide to have more kids in the future." The conversation ends, and Manager Smith does an Internet search on Down Syndrome and its genetic basis. He has concerns that Jane's child and potential future children will not only drain his company's insurance plan because of the medical care they will require but will reduce Jane's productivity and morale while increasing her absenteeism. In Scenario One, Manager Smith's conduct has clearly violated GINA.[155] He not only followed up the inadvertent disclosure with more probing questions that he should have known would reveal genetic information but also followed up the conversation with Internet research aimed at uncovering more information about the genetic basis of Down Syndrome.[156]

In Scenario Two, Manager Smith asks Employee Jane, "How are you doing today?" Jane responds, "I'm not doing very well today. We just got the amniocentesis results back, and there is something wrong with the baby." Manager Smith responds, "I'm sorry to hear that, Jane." By dropping the issue and not further investigating the matter, Manager Smith has avoided violating GINA.[157]

Another exception applies when a manager, supervisor, union representative, or employment agency representative acquires unsolicited genetic information about an individual or a family member. For instance, an employer has not unlawfully acquired information under GINA when an employee emails his supervisor asking permission to miss work that day because he needs to take his father to a doctor's appointment for treatment related to his Huntington's disease.[158]

Perhaps not surprisingly, the regulations accompanying GINA also take into account the interplay between social media and the inadvertent acquisition of genetic information. A covered entity is not in violation of GINA where a manager, supervisor, or union or employment agency

154. *Id.* § 1635.8(b)(1)(ii)(B).

155. *See supra* notes 153–154 and accompanying text.

156. *See supra* note 154 and accompanying text.

157. *See supra* note 153 and accompanying text.

158. 29 C.F.R. § 1635.8(b)(1)(ii)(C).

representative inadvertently acquires genetic information from a social media platform to which the profile creator granted access.[159]

Consider the following example of the interplay between social media and GINA. Joe is Ruth's manager and Facebook friend. Ruth posts the following status update on her Facebook page: "Worst day ever. My twin sister tested positive for the breast cancer gene. Prayers needed." When Joe logs on to check his messages, he sees the status update, which contains genetic information within the meaning of GINA.[160] Because Ruth accepted Joe as her Facebook friend, thus giving him permission to see her status updates, and because he did not actively search for the genetic information, the acquisition is inadvertent and not illegal.[161] There is no GINA violation here so long as Joe does not use the information to make employment decisions or follow up with more probing questions that are likely to reveal additional genetic information, such as, "Ruth, have you been tested for that gene yet?"[162]

Turning to post-offer employment-related medical examinations, "a covered entity must tell health care providers not to collect genetic information, including family medical history, as part of a medical examination intended to determine the ability to perform a job, and must take additional reasonable measures within its control if it learns that genetic information is being requested or required."[163] GINA appears to bar employers from obtaining any genetic information, including family medical information, from post-offer applicants or from examinations that are required to determine whether employees remain fit for duty.[164] However, covered entities may request or require a family medical history where necessary to comply with family and medical leave laws, to conduct a DNA analysis for law enforcement purposes, or to identify human remains.[165]

Another exception exists when all of the following conditions are met: a covered entity offers health or genetic services to a covered individual (including a participant in a wellness program); the employee provides prior, knowing, voluntary, and written authorization; only the individual or the family member receiving the services and the healthcare professional receive individually identifiable information concerning the results of the genetic services; and the identifiable information is not disclosed, except in

159. *Id.* § 1635.8(b)(1)(ii)(D).

160. *See supra* notes 109–112 and accompanying text.

161. *See supra* note 158 and accompanying text.

162. *See supra* note 154 and accompanying text.

163. 29 C.F.R. § 1635.8(d).

164. *See id.*; Regulations Under the Genetic Information Nondiscrimination Act of 2008, 74 Fed. Reg. 9056, 9061 (proposed Mar. 2, 2009) (to be codified at 29 C.F.R. pt. 1635).

165. 42 U.S.C. § 2000ff-1(b)(3), (b)(6) (Supp. III 2009).

an aggregate form that will not identify the employee.[166] Furthermore, no adverse use may be made with the information.[167]

GINA does not permit employer-sponsored voluntary wellness programs to mandate the provision of genetic information or to penalize individuals who refuse to provide it.[168] Nor may a covered entity offer a financial inducement for the provision of genetic information.[169] However, it is permissible to offer financial inducements for the completion of health risk assessments that include questions about family medical history or other genetic information, so long as the covered entity clarifies that the inducement is available whether or not the participant answers questions regarding genetic information.[170] Furthermore, voluntary wellness programs must be offered to persons with current health conditions and those whose lifestyle choices increase their likelihood of developing a condition.[171]

GINA also permits covered entities to request or require information to genetically monitor the biological effects of toxic substances in the workplace as long as the covered entity gives written notice of the monitoring to the covered individuals and, if the monitoring is not required by law, obtains the person's prior, knowing, voluntary, and written authorization.[172] GINA regulations define "genetic monitoring" as the "periodic examination of employees to evaluate acquired modifications to their genetic material, such as chromosomal damage or evidence of increased occurrence of mutations, caused by the toxic substances they use or are exposed to in performing their jobs, in order to identify, evaluate, and respond to the effects of, or to control adverse environmental exposures in the workplace."[173] The covered entity must inform the covered individual of the results of the monitoring and ensure that the monitoring complies with applicable laws and regulations.[174] The covered entity may only receive the monitoring results in an aggregate form that does not identify the monitored individual.[175]

No matter how genetic information is acquired, covered entities may not use it to make employment decisions or to limit, segregate, or classify covered persons.[176] Such information must also be stored and disclosed in compliance with GINA.[177]

---

166. 42 U.S.C. § 2000ff-1(b)(2)(A)–(D).

167. *See infra* note 176 and accompanying text.

168. 29 C.F.R. § 1635.8(b)(2)(i)(A).

169. *Id.* § 1635.8(b)(2)(ii).

170. *Id.*

171. *Id.* § 1635.8(b)(2)(iii).

172. *Id.* § 1635.8(b)(5).

173. *Id.* § 1635.3(d).

174. *Id.*

175. *Id.* § 1635.8(b)(5)(iii).

176. *See* 42 U.S.C. § 2000ff-1(c) (Supp. III 2009); Regulations under the Genetic Information Nondiscrimination Act of 2008, 74 Fed. Reg. 9056, 9061 (proposed Mar. 2,

With regard to genetic privacy, a covered entity must maintain genetic information on a separate form in a separate medical file marked "confidential" and must treat the information as a confidential medical record.[178] According to the GINA regulations, GINA does not require the removal of genetic information that was placed in personnel files prior to Title II's effective date, November 21, 2009.[179] However, GINA's prohibition on the use and disclosure of genetic information applies to all genetic information that meets the statutory definition, including information requested, required, or purchased prior to November 21, 2009.[180]

Genetic information may only be disclosed in limited circumstances: (i) to the individual upon written request; (ii) to an occupational or other health researcher "if the research is conducted in compliance with the regulations and protections provided for under part 46 of title 45, Code of Federal Regulations;" (iii) to a court in response to a judicial order; (iv) to a government official investigating GINA compliance; (v) to an entity or person checking an employee's compliance with state or federal family and medical leave laws; and (vi) to a federal, state, or local public health agency with regard to information that concerns a contagious, deadly disease or life-threatening illness if the employee is notified of the disclosure.[181] Turning to GINA's relationship to other laws protecting private health information, Title II does not prohibit covered entities from using or disclosing health information that is authorized under HIPAA.[182]

Title II permits disparate treatment but not disparate impact claims.[183] Nothing in Title II may be construed to provide for the enforcement of or penalties for Title I violations,[184] and Title II does not apply retroactively.[185] GINA's enforcement mechanisms mirror those of Title VII of the Civil Rights Act of 1964 ("Title VII").[186] Before filing an action in court, a

---

2009) (to be codified at 29 C.F.R. pt. 1635) ("Covered entities are cautioned, however, that the use of genetic information to discriminate, no matter how that information may have been acquired, is prohibited.").

177. 42 U.S.C. § 2000ff-1(c).

178. *Id.* § 2000ff-5(a); 29 C.F.R. § 1635.9.

179. 29 C.F.R. § 1635.9(a)(5).

180. *Id.*

181. 42 U.S.C. § 2000ff-5(b).

182. *Id.* § 2000ff-5(c).

183. *Id.* § 2000ff-7(a) ("'[D]isparate impact' . . . on the basis of genetic information does not establish a cause of action under this Act.").

184. *Id.* § 2000ff-8(a)(2)(B)(i).

185. *See* Citron v. Niche Media/Ocean Drive Magazine, No. 10-24014-CIV, 2011 U.S. Dist. LEXIS 13070, at *5–*6, 2011 WL 381939, at *2 (S.D. Fla. Feb. 2, 2011) (explaining that purported GINA violations that occurred prior to GINA's effective date of November 21, 2009, are not actionable).

186. 42 U.S.C. § 2000ff-6(a).

person must file a complaint against the employer with the EEOC.[187] Like Title VII, GINA requires claimants to exhaust administrative remedies prior to filing suit in court and caps damages available against employers depending on employer size.[188] GINA does not preempt more protective state or local laws that address genetic discrimination in employment or privacy protection for genetic information.[189] Likewise, GINA does not impact a person's rights under other anti-discrimination laws, such as the ADA.[190]

A. Justifying GINA

Two primary criticisms have been lodged against Title II of GINA. According to the first line of criticism, GINA is unnecessary because genetic discrimination in employment does not exist, and even if it does, the problem is not so significant as to warrant legislative intervention.[191] To the extent such intervention is warranted, pre-existing legislation, including the ADA and state anti-discrimination laws, adequately addresses it, rendering GINA unnecessary.[192]

Under the second line of criticism, even if genetic testing in employment may give rise to discrimination and related concerns, it should be permitted, at least under some circumstances, because employers may use genetic information for laudable purposes.

B. Is GINA Necessary?

Turning to the first criticism, opponents of GINA argue that it is unnecessary because of the lack of evidence to support the belief that employers discriminate on the basis of genetic information.[193] The Genetic

---

187. *See* U.S. EQUAL EMPLOYMENT OPPORTUNITY COMMISSION, *Filing a Charge of Discrimination*, http://www.eeoc.gov/employees/charge.cfm.

188. *See* Slaughter, *supra* note 10, at 80 (discussing the enforcement and damages provisions of GINA before its enactment).

189. Genetic Information Nondiscrimination Act of 2008, 29 C.F.R. § 1635.11 (2011); Regulations under the Genetic Information Nondiscrimination Act of 2008, 74 Fed. Reg. 9056, 9064 (proposed Mar. 2, 2009) (to be codified at 29 C.F.R. pt. 1635).

190. Regulations under the Genetic Information Nondiscrimination Act of 2008, 74 Fed. Reg. at 9064 ("GINA does not affect an individual's rights under the ADA.").

191. *See infra* notes 194–98 and accompanying text.

192. *See infra* notes 224–28 and accompanying text.

193. *See* Henry T. Greely, *Genotype Discrimination: The Complex Case for Some Legislative Protection*, 149 U. PA. L. REV. 1483, 1483 (2001) ("Genetic discrimination is a much greater threat in people's fears than it is in reality, today or in the foreseeable future, for both scientific and social reasons."); Ann Scott Tyson, *Lawmakers Play Catch-Up to Genetic Science*, CHRISTIAN SCI. MONITOR, Aug. 10, 2000, at 3, *available at* http://www.csmonitor.com/2000/0810/p3s1.html ("Today, genetic discrimination remains relatively rare. Fewer than 1 percent of employers conduct genetic testing, according to a

Information Nondiscrimination in Employment Coalition[194] widely criticized an earlier version of GINA as unnecessary, as provoking frivolous lawsuits, and as containing an overbroad definition of genetic information.[195]

Some economists argue that legislative regulation was unnecessary because market forces would have punished employers if genetic discrimination is as irrational as its critics claim. The Labor Market Hypothesis suggests that employers who irrationally discriminate against productive workers will eventually be "punished by market forces," compelling them to stop such irrational behavior.[196] According to this theory, if genetic discrimination is entirely irrational, GINA is unnecessary because economic forces alone will rein in discriminatory behavior.[197]

This economics-based argument fails because it assumes a competitive labor market with employers and workers who have perfect information and no barriers to job mobility. However, we do not live in a perfect world. Labor markets in high unemployment areas are not competitive, and many people have barriers to job mobility, such as persons who lack high school diplomas or college degrees or individuals who have developed employer-specific skills making them less attractive to other employers. Furthermore, neither employers nor employees ever have perfect information as evidenced by the aforementioned assertions that many employers have a misplaced belief in the notion of genetic determinism.[198]

More importantly, at least some evidence indicates that genetic discrimination in employment was a problem prior to GINA's enactment. For instance, the United States Council for Responsible Genetics reported 200 cases of genetic discrimination in employment.[199] A 2004 study revealed that 92% of participants did not want their employers to have

---

2000 survey by the American Management Association."); *see also* Cheye Calvo, *Engineering Genetics Policy*, STATE LEGISLATURES, Sept. 2000, at 28, 29 (When asked if he believed whether genetic discrimination currently existed, Senator Patrick Johnston, who has already sponsored at least six genetics bills in California, admitted, "I hope it doesn't, but I don't know."). *But see* Andrews, *supra* note 13, at 47 (A 1989 survey by the Office of Technology Assessment reported that one in twenty companies used genetic tests to screen and monitor workers.).

194. Slaughter, *supra* note 10, at 78.

195. Sismondo, *supra* note 96, at 473–74, 473 n.84.

196. Colin S. Diver & Jane Maslow Cohen, *Genophobia: What Is Wrong with Genetic Discrimination?*, 149 U. PA. L. REV. 1439, 1464 (2001).

197. *See id.*

198. *See supra* notes 54, 94 and accompanying text.

199. Louise Reohr, *Got an Illness in the Family? In That Case, You're Fired*, THE INDEPENDENT (London), Feb. 6, 2001, at 10; *see also* Marisa Anne Pagnatarro, *Genetic Discrimination and the Workplace: Employee's Right to Privacy v. Employer's Need to Know*, 39 AM. BUS. L.J. 139, 154–55 (2001) (discussing several surveys and circumstances revealing cases of genetic discrimination).

access to their genetic information,[200] and a 2000 Northwestern National Life Survey revealed that 15% of employers intended to obtain applicants' genetic information before extending final employment offers to them.[201] Of the 332 participants in a Georgetown survey, 13% believed that they were denied or terminated from employment due to genetic discrimination.[202] A 1996 study documented over 200 cases of asymptomatic carriers of genetic disorders or individuals predisposed for genetic disorders who reported various discriminatory actions by employers and insurers.[203] Another survey of health care providers revealed that respondents knew of 582 instances in which asymptomatic employees had been refused employment or insurance based on their genetic predispositions.[204]

Surveys conducted prior to GINA's enactment suggest that fear of genetic discrimination chilled people's willingness to undergo genetic testing, even when testing could permit them to take prophylactic measures.[205] A Georgetown study discovered that nearly one in ten individuals refused to undergo genetic testing for fear of discrimination.[206] Two-thirds of the respondents to a 1997 survey refused to undergo genetic testing if employers and insurers could view the results, and 85% felt that employers should be forbidden from obtaining genetic information about their employees.[207] In a 2002 Harris Poll, 81% of participants reported that it was a "good thing" to be able to use genetic testing to discern people's predispositions to diseases.[208] Over 85% of the respondents in a 1995 poll were either "very" or "somewhat" concerned about employer access to and misuse of genetic information.[209]

A survey of physicians offering genetic testing for breast cancer revealed that 70% had had at least one patient who refused the test, "68% of

---

200. Slaughter, *supra* note 10, at 71.

201. Ashley M. Ellis, Comment, *Genetic Justice: Discrimination by Employers and Insurance Companies Based on Predictive Genetic Information*, 34 TEX. TECH L. REV. 1071, 1082 (2003).

202. Bryce A. Lenox, Comment, *Genetic Discrimination in Insurance and Employment: Spoiled Fruits of the Human Genome Project*, 23 U. DAYTON L. REV. 189, 194 (1997); *see also* Torben Spaak, *Genetic Discrimination*, 7 MINN. J. L. SCI. & TECH. 639, 642–45 (2006) (discussing case studies revealing potential instances of genetic discrimination in employment and insurance).

203. Miller, *supra* note 50, at 190.

204. *Id.*

205. Aaron Zitner, *Senate Close To Passing Bill Banning Genetic Discrimination*, L.A. TIMES, May 22, 2003, at A30; *see* 2 ENCYCLOPEDIA OF BIOETHICS, *supra* note 27, at 956; *Genetic Testing Raises Fears of Workplace Bias*, *supra* note 20.

206. Miller, *supra* note 50, at 190.

207. *Id.* at 189.

208. Meera Adya & Brian H. Bornstein, *Genetic Information and Discrimination in Employment: A Psycho-Legal Perspective*, 32 WM. MITCHELL L. REV. 265, 277–78 (2005).

209. Miller, *supra* note 50, at 189.

patients declining the test did so out of fear that confidentiality would not be maintained, [and] 52% feared the actual results of the test."[210] People are so concerned about keeping their test results private that many pay for the testing themselves and even use "shadow" charts and aliases so that the results will never be included in their medical, insurance, and personnel records.[211] Some researchers complain that people's refusals to undergo genetic testing make long-term research studies virtually impossible to complete.[212]

On the other hand, as of 2006, no case of genetic discrimination in employment had been *decided* by a United States court.[213] Since Title II of GINA took effect, only a handful of cases alleging GINA violations have been filed, mostly by *pro se* litigants. Because GINA is still in its infancy, the present dearth of GINA litigation does not indicate that genetic discrimination did not require redress or that GINA was unnecessary.

A landmark case that predated GINA explains why. In 2001, the EEOC filed suit against Burlington Northern & Santa Fe Railway ("BNSF"), alleging genetic discrimination in employment.[214] The EEOC claimed that BNSF had forced employees who filed ADA claims alleging work-related Carpal Tunnel Syndrome to undergo genetic testing for a genetic marker on Chromosome 17[215] that indicates Hereditary Neuropathy with Liability to Pressure Palsies, which is often linked to the development of Carpal Tunnel Syndrome.[216] Purportedly the test's purpose was to determine each employee's genetic susceptibility to Carpal Tunnel Syndrome in order to mitigate BNSF's ADA and worker's compensation liability.[217] According to the complaint, BNSF neither informed employees of the test's purpose nor obtained their informed consent.[218] An employee who refused to take the test was allegedly "threatened with imminent discharge."[219] The EEOC

---

210. Adya & Bornstein, *supra* note 208, at 279.

211. Melanie Payne, *Genetic Fears*, CHI. TRIB., Apr. 29, 1998, at 7, *available at* http://articles.chicagotribune.com/1998-04-29/news/9804300181_1_genetic-diseases-genetic-counselor-genetic-disorders.

212. For a different perspective, *see* Greely, *supra* note 193, at 1498–99, which argues that proponents of genetic discrimination legislation intentionally scared people about the likely significance of current and future genetic discoveries in order to build support for legislation.

213. Sismondo, *supra* note 96, at 474. For a different case in which workers alleged that their employer discriminated against them due to their genetic predispositions to develop Carpal Tunnel Syndrome, *see EEOC v. Woodbridge Corp.*, 263 F.3d 812 (8th Cir. 2001).

214. Sismondo, *supra* note 96, at 475 n.92.

215. Karen Long, *EEOC Leader to Speak on Genetic Discrimination*, CLEVELAND PLAIN DEALER, Sept. 12, 2001, at E1.

216. Sismondo, *supra* note 96, at 475.

217. *Id.*

218. *Id.* at 476.

219. *Id.* (internal quotation marks omitted).

sought relief under the ADA and demanded a preliminary injunction against BNSF to prohibit both the genetic testing of employees who filed disability claims and any adverse action or termination against employees who refused to undergo the testing.[220] The case settled in April 2001.[221] As part of the settlement, BNSF agreed to halt its genetic testing and to pay $2.2 million in damages to affected employees.[222] GINA removes any doubt as to whether such conduct is lawful and aims to either prevent such misconduct or, at the very least, to provide employees with legal redress when it does occur.

C. Did Pre-Existing Law Adequately Protect Against Genetic Discrimination?

Critics further claim that GINA is redundant because existing law adequately protected against genetic discrimination.[223] They argue that, taken together, federal statutes, state and local anti-discrimination or human rights laws, and even the tort of "wrongful discharge in violation of public policy" provided sufficient protection against genetic discrimination in employment and obviated the need to squander valuable time and resources to pass a federal law that may prove difficult and costly for covered entities to implement and that may clog courts with frivolous litigation.[224]

However, the Advisory Committee on Genetics, Health, and Society at the Department of Health and Human Services came to a different conclusion, opining that current federal law "does not adequately protect against discrimination based on genetic predisposition."[225] Likewise, in June 2005, the Genetic Discrimination Task Force stated that existing law and judicial decisions left "substantial gaps" in coverage and failed to provide necessary safeguards against genetic discrimination in employment.[226] The drafters of GINA agreed, observing in the statute's findings that "[f]ederal law addressing genetic discrimination in health

---

220. *Id.* at 475–76.

221. *Id.* at 476.

222. *Id.* at 477.

223. Katherine Hathaway, Comment, *Federal Genetic Nondiscrimination Legislation: The New "Right" and the Race to Protect DNA at the Local, State, and Federal Level*, 52 CATH. U. L. REV. 133, 157 (2002).

224. *See id.*; Candice Hoke, *Reasons To Eschew Federal Lawmaking and Embrace Common Law Approaches to Genetic Discrimination*, 16 J.L. & HEALTH 53, 56–58 (2002) (recommending the use of Ohio's tort of wrongful discharge in violation of public policy as a model to provide legal recourse for genetic discrimination in employment).

225. Slaughter, *supra* note 10, at 73 (2006) (internal quotation marks omitted). The Committee sent Secretary Michael Leavitt a letter urging him to press Congress to enact federal legislation to address the problem of genetic discrimination. *Id.*

226. *Id.*

insurance and employment is incomplete in both the scope and depth of its protections."[227]

D. ADA

In 1990, Congress enacted the ADA, which prohibits discrimination against "an individual who, with or without reasonable accommodation, can perform the essential functions of the employment position that such individual holds or desires."[228] Although the ADA does not explicitly address or define genetic discrimination, "[i]n 1995, the EEOC adopted the view that the ADA prohibits discrimination against workers based on their genetic makeup."[229]

A "disability" under the ADA is a physical or mental impairment that presently and substantially limits the person in one or more major life activities, a record of such impairment or being "regarded as" having such an impairment by an employer.[230] Under the pre-amended ADA, disability was determined with regard to any measures the individual had taken to correct or mitigate the alleged disability.[231] The decision in *Sutton v. United Air Lines* indicated that courts must look to a person's actual condition in order to determine whether a person has a disability as defined by statute.[232] The ADA forbids an employer from considering a qualified individual's disability for purposes of recruitment, hiring, job application procedures, promotion, tenure, demotion, transfer, lay-off, termination, return from lay-

---

227. Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-233, § 2(5), 122 Stat. 881, 882 (2008) (codified as amended at 42 U.S.C. §§ 2000ff to 2000ff-12 (Supp. III 2009)).

228. Americans with Disabilities Act of 1990, 42 U.S.C. § 12111(8) (2006).

229. Paul Miller, *Analyzing Genetic Discrimination in the Workplace*, HUM. GENOME NEWS (U.S. Department of Energy Office of Biological and Environmental Research, Oak Ridge, Tenn.), Feb. 2002, at 9, *available at* http://www.ornl.gov/sci/techresources/Human_Genome/publicat/hgn/v12n1/09workplace.shtml ("Though lacking the force of law, the EEOC's policy explicitly states that discrimination on the basis of genetic information is covered under the third prong of the statutory definition of 'disability,' which covers people who are regarded as having impairments."). Although the ADA was recently amended, this Article refers to the version of the ADA in existence prior to GINA and the case law arising therefrom, unless otherwise indicated.

230. 42 U.S.C. § 12102.

231. *See* Sutton v. United Air Lines, Inc., 527 U.S. 471 (1999). The Court overruled the EEOC Guidelines and held that disability must be considered with respect to mitigating measures so that the plaintiffs who suffered from severe myopia were not presently disabled by their visual condition, which was mitigated with the use of glasses and contact lenses that provided them with near 20/20 vision. *Id.* Notably, the Americans with Disabilities Amendments Act overruled *Sutton*, though contact lenses may still be taken into account when determining whether a person has a disability as defined under the ADA. 42 U.S.C. § 12102(4)(E).

232. *See Sutton*, 527 U.S. at 488–89.

off, rehiring, rate of pay, job assignment or classification, leave of absence, sick leave, fringe benefits, seniority lists, training, and employer-sponsored activities, among other things.[233]

To establish a prima facie case of discrimination under the ADA, an individual must prove each of the following factors by a preponderance of the evidence: (i) the employer has fifteen or more employees and otherwise satisfies the requirements to be considered a covered entity; (ii) the claimant is a person with a disability, has a record of disability, or is regarded as a person with a disability within the meaning of the ADA; (iii) the claimant is otherwise qualified to perform the essential functions of the job, with or without reasonable accommodation; and (iv) the claimant suffered an adverse employment action due to the alleged disability.[234]

Critics of GINA argue that a person may sue for genetic discrimination in employment under the ADA by proving that the employer mistakenly regards the individual as having a disability. However, as the subsequent discussion illuminates, the pre-amended ADA failed to sufficiently protect against genetic discrimination in employment because, *inter alia*, the EEOC guidelines, which indicated that the ADA forbids genetic discrimination, lacked the force of law. Further, the ADA permitted pre- and post-employment genetic testing and did not prohibit the request or disclosure of genetic information.

E. EEOC Guidelines

In March 1995, the EEOC issued guidelines indicating that the ADA defined disability as including "individuals who are subjected to discrimination on the basis of genetic information relating to illness, disease or other disorders."[235] In fact, certain diseases with genetic components, such as multiple sclerosis and muscular dystrophy, are classified as disabilities under the ADA.[236] However, EEOC guidelines lack the force of law and can be overruled if a court finds them to be arbitrary, capricious, an abuse of discretion, or manifestly contrary to law.[237] Thus, unlike GINA, the EEOC guidelines offered less certainty of protection against genetic discrimination.

---

233. Kirke D. Weaver, *Genetic Screening and the Right Not To Know*, 13 ISSUES IN L. & MED. 243, 266 (1997).

234. Pagnatarro, *supra* note 199, at 159; *see* 42 U.S.C. §§ 12102(1), 12111(5), 12111(8), 12112(a).

235. Pagnatarro, *supra* note 199, at 162; Lenox, *supra* note 202, at 205 (internal quotation marks omitted).

236. Weaver, *supra* note 233, at 262.

237. Miller, *supra* note 50, at 191.

F. Pre- and Post-Employment Genetic Testing

Prior to GINA, the ADA appears to have permitted employers to withdraw conditional employment offers from applicants whose mandatory pre-employment genetic tests revealed "bad" genes so long as employers subjected all prospective employees to the same pre-entrance exams.[238] Nor did the ADA restrict the exam's underlying purpose.[239] In fact, EEOC guidelines stated that "medical examinations [of applicants] . . . do not have to be job-related and consistent with business necessity."[240]

By way of illustration, in *Norman-Bloodsaw v. Lawrence Berkeley Laboratory*, present and former employees alleged that during the course of mandatory pre-employment entrance exams, their employer tested their blood and urine for syphilis, pregnancy, and sickle cell anemia without their knowledge or consent.[241] Because African-American and female employees were singled out for testing, the court upheld the employees' Title VII claims.[242] However, the Ninth Circuit dismissed the ADA claims because "[t]he ADA poses no restriction on the scope of the entrance examinations; it only guarantees the confidentiality of the information gathered and restricts the use to which an employer may put the information."[243] The holding signifies that under the ADA, pre-employment entrance exams do not have to be job-related and consistent with business necessity.[244] Consequently, the ADA inadequately protected against genetic discrimination in employment; in contrast, GINA expressly forbids pre- and post-employment genetic testing, except under very limited circumstances.

Furthermore, prior to the enactment of GINA and the ADA Amendments Act, an employer could genetically test current employees if the test was job-related and consistent with business necessity under Section 12113(a) of the ADA.[245] Some courts liberally construed that standard. For example, in *Milton v. Scrivner, Inc.*, the Tenth Circuit held that a qualification standard that required grocery warehouse workers to move items more quickly was permissible because it was job-related and

---

238. *Id.*; Nicole Silvestri, Comment, Echazabal *and the Threat to Self-Defense: The Most Recent Call for a Consistent, Interstate Genetic Nondiscrimination Policy*, 7 U. PA. J. LAB. & EMP. L. 409, 417 (2005). Notably, however, medical examinations of current employees must be "job-related and consistent with business necessity." Genetic Information Nondiscrimination Act of 2008, 29 C.F.R. § 1630.14(b)(3) (2011).

239. Melinda B. Kaufmann, *Genetic Discrimination in the Workplace: An Overview of Existing Protections*, 30 LOY. U. CHI. L.J. 393, 407 (1999).

240. 29 C.F.R. §1630.14(b)(3).

241. Norman-Bloodsaw v. Lawrence Berkeley Lab., 135 F.3d 1260, 1264–65 (9th Cir. 1998); *see* Ellis, *supra* note 201, at 1083–84.

242. *Norman-Bloodsaw*, 135 F.3d at 1271–73; Ellis, *supra* note 201, at 1084.

243. *Norman-Bloodsaw*, 135 F.3d at 1273; Ellis, *supra* note 201, at 1085.

244. *See Norman-Bloodsaw*, 135 F.3d at 1273; Ellis, *supra* note 201, at 1085

245. Silvestri, *supra* note 238, at 418.

consistent with business necessity.[246] This held true even though the standard screened out many applicants with disabilities.[247] The standard was implemented to improve employee efficiency and productivity and to increase the employer's competitiveness in the market.[248] Before GINA, an employer could perhaps have justified its genetic monitoring and screening policies because "injured" or "unhealthy" employees arguably decrease efficiency, productivity, and market competitiveness. However, GINA strictly prohibits covered entities from requiring such testing and from basing employment decisions on the results of such tests.

G. Defining "Disability" After Bragdon

GINA critics also erroneously argued that GINA is redundant because a landmark disability case—*Bragdon v. Abbott*[249]—proved that the ADA applies with equal force to asymptomatic conditions, even genetic conditions, predispositions, and carrier status. However, the ADA covers manifested medical conditions, as opposed to genetic predispositions or carrier status and, furthermore, provides little or no genetic privacy protections.[250]

Critics of GINA claim that *Bragdon* established that genetic discrimination fell within the scope of the ADA; however, such a reading is unavailing. In *Bragdon*, an asymptomatic HIV-infected woman sought to have a cavity filled.[251] Upon learning that she was HIV-infected, her dentist told her that he would only fill her cavity if she rented and paid for the use of hospital facilities.[252] The dentist feared that treating her in his office using his equipment would pose a "direct threat" to his other patients.[253] She subsequently sued the dentist under Title III of the ADA, which, *inter alia*, prohibits private providers of public accommodations from discriminating against patrons on the basis of disability.[254] The district court granted her motion for summary judgment, and the First Circuit affirmed.[255] The United States Supreme Court granted a petition for certiorari to determine whether an asymptomatic HIV-infected individual is a person with a disability for purposes of the ADA.[256] The Court held that Ms. Abbott was a person with a disability because she had a physical

246. Milton v. Srivner, Inc., 53 F.3d 1118, 1124 (10th Cir. 1995).
247. *See id.*
248. *Id.*
249. 524 U.S. 624 (1998).
250. *See* Americans with Disabilities Act of 1990, 42 U.S.C. §§ 12101–12213 (2006).
251. *Bragdon*, 524 U.S. at 628–29.
252. *Id.* at 629.
253. *Id.* at 629–30.
254. *Id.* at 629; *see* 42 U.S.C. § 12182.
255. *Bragdon*, 524 U.S. at 629–30.
256. *Id.* at 628.

impairment that substantially limited her in the major life activity of reproduction.[257] However, the dissent pointed out that Ms. Abbott's decision not to reproduce was not a substantial limitation because it is a voluntary choice.[258]

GINA critics relied upon *Bragdon* to argue that the ADA adequately protects individuals against genetic discrimination in employment, making GINA unnecessary. This argument fails for several reasons. First, according to Justice Anthony Kennedy, "HIV is not included in the list of specific disorders constituting physical impairments, in part because HIV was not identified as the cause of AIDS until 1983. . . . HIV infection does fall well within the general definition set for by the regulations, however."[259] Scientists knew that genes, unlike the causative agent of AIDS, were the cause of genetic disorders long before the ADA was enacted. Therefore, if Congress had intended the ADA to cover discrimination on the basis of a person's genotype, it could easily have drafted the ADA to include words to that effect. In fact, after *Bragdon*, HIV and AIDS were added to the list of diseases considered to be disabilities under the ADA.[260] No such change was ever made to indicate that discrimination on the basis of a person's genotype or asymptomatic carrier status violates the ADA.

Expanding the ADA to embrace all asymptomatic persons with genetic maladies, such as a person who carries an as-of-yet unexpressed dominant allele for Huntington's disease, would have arguably violated the legislative intent underlying the ADA because Congress clearly intended to "restrict its protections and requirements to include only a confined and *historically disadvantaged* group . . . [those] 'with basically phenotypic disabilities who had a long history of discrimination.'"[261] Moreover, during legislative hearings, Congress referenced only those persons who were both presently disabled and legally disabled under the ADA.[262] Because Congress never discussed whether the ADA should cover individuals who had only the potential of developing a disability because of "current genetic misspellings," the only way to find that the ADA embraced these asymptomatic individuals with "bad" genes is "through the negative inference that the failure to mention them indicates only that the ADA does not specifically *exclude* them."[263]

As Justice Kennedy articulated:

---

257. *Id.* at 631.

258. *Id.* at 661 (Rehnquist, J., dissenting).

259. *Id.* at 633.

260. *See* Nondiscrimination on the Basis of Disability in State and Local Government Services, 28 C.F.R. § 35.104(1)(ii) (2010).

261. Kimberly A. Steinforth, Note, *Bringing Your DNA to Work: Employers' Use of Genetic Testing under the Americans with Disabilities Act*, 43 ARIZ. L. REV. 965, 994 (2001).

262. *Id.* at 995.

263. *Id.*

> The disease [AIDS] follows a predictable and, as of today, unalterable course. Once a person is infected with HIV, the virus invades different cells in the blood and in body tissues. . . . In light of the immediacy with which the virus begins to damage the infected person's white blood cells and the severity of the disease, we hold it is an impairment from the moment of infection. As noted earlier, infection with HIV causes immediate abnormalities in a person's blood, and the infected person's white cell count continues to drop throughout the course of the disease, even when the attack is concentrated in the lymph nodes. In light of these facts, HIV infection must be regarded as a physiological disorder with a constant and detrimental effect on the infected person's hemic and lymphatic systems from the moment of infection. HIV infection satisfies the statutory and regulatory definition of a physical impairment during every stage of the disease.[264]

Notably, HIV is a retrovirus rather than a genetic condition.[265] Unless expressed, a "bad" genotype does not cause "immediate abnormalities" like AIDS.[266] For a carrier or a person whose genetic predisposition is never triggered, his or her bad gene(s) will never cause abnormalities. Many carrier genes and genetic predispositions cause no immediate and severe damage to the carrier's bodily systems. In contrast, an asymptomatic HIV-infected person has not yet manifested symptoms because the virus, which is already affecting the victim's immune system and bodily functions, has not impacted that individual to the threshold point at which a doctor gives a diagnosis of full-blown AIDS.[267] However, as Justice Kennedy stated, HIV has a "constant and detrimental effect."[268] Clearly then, a strong argument exists that many genotypes that could give rise to genetic discrimination in employment do not fall within the general definition of disability under the ADA, even as interpreted by the *Bragdon* Court.

Policy dictates discouraging an HIV-infected person from spreading such a highly contagious disease via unprotected sex or other means.[269] In contrast, bad genes are not contagious or sexually transmitted. Although unprotected sex poses a risk of transmitting a bad gene to potential offspring borne as a result of intercourse, a carrier's bad gene is not sexually transmitted to sexual partners or through the exchange of bodily fluids; thus, a carrier may have intercourse, even unprotected, without fear of infecting his partner with the bad allele.[270] Therefore, the individual's

---

264. Bragdon v. Abbott, 524 U.S. 624, 634, 637 (1998).

265. *Id.* at 634.

266. *See supra* notes 25–33 and accompanying text (discussing genotypes and phenotypes).

267. *Bragdon*, 524 U.S. at 637.

268. *Id.*

269. *Id.* at 639, 640.

270. Laura F. Rothstein, *Genetic Discrimination: Why* Bragdon *Does Not Ensure Protection*, 3 J. HEALTH CARE L. & POL'Y 330, 344–45 (2000).

limitations for having intimate relationships are not nearly as substantial as Ms. Abbott's. Genetic carriers may safely have intimate partners, and with the use of birth control and other contraceptive methods, they may significantly reduce the risk of pregnancy. Considering the inheritance patterns of various diseases and the likelihood of passing along a bad allele, the chance of producing a child suffering an autosomal recessive genetic disorder is 0.25% in contrast to the 25% risk of transmission to both partner and child that is inherent in sex with an HIV-infected person according to some source.[271]

Carriers of recessive alleles have a 50% risk of passing on their allele, but there is no risk of their child manifesting that allele.[272] For example, if Joe carries a recessive allele for PKU, the only way that he has any chance of having a child with PKU is if he marries another carrier of the recessive allele, which is highly improbable. Even then the chance of having a child who suffers PKU is only one in four. Furthermore, passing on the gene can be prevented by pre-childbearing genetic screening. Therefore, Joe would not have to abstain from procreation.

Individuals with dominant, codominant, or sex-linked alleles (whether recessive or dominant), which have higher rates of expression, are likelier to choose not to reproduce.[273] However, it seems intuitively unfair that under the ADA, some carriers would be classified as persons with disabilities entitled to legal protection while others would not merely because of the nature of their allele's inheritance pattern—something that is entirely beyond their control.[274] Furthermore, most people become aware of their genetic condition after having children. These parties likely could not argue that they are substantially limited in the major life activity of reproduction.[275]

Lower courts have interpreted *Bragdon* in different ways. In *Berk v. Bates Advertising USA, Inc.*, the court held that breast cancer substantially limited the major life activity of reproduction by making pregnancy "unduly risky" because the cancer treatment included a medical recommendation to have a prophylactic ovariectomy or hysterectomy.[276] In *Cornman v. N.P. Dodge Management Co.*, the court held that because a woman's breasts are an "integral part of her sexuality," a mastectomy to treat breast cancer would adversely impact her ability to have intercourse

---

271. *See Bragdon*, 524 U.S. at 639–40; Rothstein, *supra* note 270, at 344. With antiretroviral therapy, the perinatal risk of transmission can sometimes be reduced to as low as 8%. *Bragdon*, 524 U.S. at 640; Rothstein, *supra* note 270, at 344.

272. *See supra* notes 27–28 and accompanying text.

273. *See* Rothstein, *supra* note 270, at 345.

274. *See* Implementation of Executive Order 12250, Nondiscrimination on the Basis of Handicap in Federally Assisted Programs, 28 C.F.R. § 41.31(b) (2010).

275. *See* Rothstein, *supra* note 270, at 346.

276. Berk v. Bates Adver. USA, Inc., 25 F. Supp. 2d 265, 268–69 (S.D.N.Y. 1998).

with others.[277] On the other hand, however, in at least two other cases, plaintiffs who relied on the inability to reproduce as proof of their disability did not prevail.[278] However, if reproduction is the only major life activity in which a victim of genetic discrimination could claim a substantial limitation, then one has to wonder whether only persons of childbearing age and ability (*i.e.*, fertile, young men and women) would have been covered under the ADA.

Another reason that *Bragdon* is arguably inapplicable in the context of genetic discrimination is that while HIV is likely to be a contributing factor in an infected individual's cause of death, having a genetic predisposition or carrier status is not. As opposed to Ms. Abbott who was likely to develop AIDS in the future, a PKU carrier will never develop the condition and a carrier of the BRCA-1 gene has a heightened risk of developing breast cancer but may never develop it. *Bragdon* does not stand for the proposition that a person predisposed to develop a serious disease is a person with a disability or a person regarded as having a disability under the ADA. In the context of HIV/AIDS, one could just as easily argue that drug users or those engaging in unprotected sexual intercourse are predisposed to developing AIDS and, therefore, should be protected under the ADA. While the latter argument seems absurd and would undoubtedly fail in court, such overreaching may have been necessary to stretch *Bragdon*'s narrow holding to embrace all forms of genetic discrimination in employment.

Perhaps most importantly, Justices Rehnquist, Scalia, and Thomas dissented in *Bragdon*, signaling how they would rule if an ADA claim alleging genetic discrimination came before them. In particular, Justice Rehnquist remarked:

---

277. Cornman v. N.P. Dodge Mgmt. Co., 43 F. Supp. 2d 1066 (D. Minn. 1999). *But see* Schwertfager v. Boynton Beach, 42 F. Supp. 2d 1347, 1359–60 (S.D. Fla. 1999) (refusing to recognize a woman's breast cancer and resulting surgery as a disability under the ADA).

278. *See* Gutwaks v. Am. Airlines, No. 3:98-CV-2120-BF, 1999 U.S. Dist. LEXIS 16833, 1999 WL 1611328 (N.D. Tex. Sept. 2, 1999); Qualls v. Lack's Stores, Inc., No. 5:98-CV149-C, 1999 U.S. Dist. LEXIS 5732, 1999 WL 731758 (N.D. Tex. Mar. 31, 1999). In *Gutwaks*, an HIV-infected man argued that he was disabled under the ADA because he could not reproduce. *Gutwaks*, 1999 WL 1611328, at *1–*2. However, the court relied on *Sutton v. United Air Lines, Inc.*, 527 U.S. 471 (1999) to argue that whether a person is disabled under the ADA hinges on an individualized assessment of whether the physical impairment substantially limits the individual claimant's major life activities. *Id.* at *4–*5. Because the plaintiff admitted that he did not have children and had never desired to have them, the court ruled that he was not disabled under the ADA. *Id.* at *5. In *Qualls*, the plaintiff claimed that his Hepatitis C made him disabled under the ADA because it prevented him from reproducing. *Qualls*, 1999 WL 731758, at *5. However, because he had already undergone a vasectomy after having children with his wife, the court concluded that his inability to reproduce was not linked to his Hepatitis C. *Id.* at *3. Therefore, he was not person with a disability for purposes of the ADA. *Id.*

> Asymptomatic HIV does not presently limit respondent's ability to perform any of the tasks necessary to bear or raise a child. Respondent's argument, taken to its logical extreme, would render every individual with a genetic marker for some debilitating disease "disabled" here and now because of some possible future effects.[279]

Furthermore, Justice Ginsburg's concurrence in *Bragdon* emphasized that HIV "has been regarded as a disease limiting life itself."[280] She appears to have been heavily influenced by the American Medical Association's brief stating that HIV "pervades life choices" and affects one's ability to obtain health care and to care for oneself.[281] Because those same concerns do not necessarily arise for a genetic carrier or an asymptomatic person, it is uncertain whether Justice Ginsburg would have changed her vote if the situation involved a case of genetic discrimination in employment. She emphasized HIV's adverse impact on the ability to have intimate relationships due in part to people's somewhat irrational fear of catching the virus.[282] However, no such concern arises in the context of all bad genes, which unlike a virus, cannot spread from one person to another merely by sex, blood transfusions, drug use, or the transmission of certain bodily fluids.

Though it involves genetic discrimination in insurance and not employment, some argue that *Katskee v. Blue Cross/Blue Shield of Nebraska* illustrates that at least some genetic conditions fall within the ADA's definition of disability.[283] In *Katskee*, the Supreme Court of Nebraska broadly defined "bodily disorders" as follows:

> [an] illness, encompass[ing] any abnormal condition of the body or its components of such a degree that in its natural progression would be expected to be problematic; a deviation from the healthy or normal state affecting the functions or tissues of the body; an inherent defect of the body; or a morbid physical or mental state which deviates from or interrupts the normal structure or function of any part, organ, or system of the body and which is manifested by a characteristic set of symptoms and signs.[284]

The court held that a woman who was asymptomatic, but possessed a predisposition for breast and ovarian cancer, did suffer from a bodily disorder or disease within the meaning of her insurance policy's terms

---

279. *Bragdon*, 524 U.S. at 661 (Rehnquist, J., concurring in judgment and dissenting in part).

280. *Id.* at 656 (Ginsburg, J., concurring).

281. *Id.*

282. *Id.* at 656–57.

283. *See* Katskee v. Blue Cross/Blue Shield of Neb., 515 N.W.2d 645 (Neb. 1994).

284. *Id.* at 651.

because her condition was a "genetic deviation" from an average woman's makeup.[285]

While *Katskee* indicates that at least some courts would consider a genetic predisposition for certain debilitating diseases to be a disability under the ADA, the case provides little guidance because it interpreted the terms of an insurance policy rather than the ADA or other federal statute. Considering a predisposition as a disability for insurance purposes is very different from treating it as such in an employment context (*e.g.*, hiring, promoting, or firing) because insurance risk classifications require insurers to speculate. For instance, not all smokers will develop lung cancer, but they are legally charged higher insurance premiums on the assumption that they are more likely to develop it than a non-smoker. Yet a policy of not hiring smokers because they are likelier to have higher absenteeism does not necessarily follow from using smoking as an insurance risk classification. Moreover, *Katskee* was decided before *Bragdon*.

H. Regarded As "Genetically Disabled"

In the alternative, critics of GINA argued that even had the ADA's first two prongs provided insufficient protection against genetic discrimination in employment, victims could have sought legal recourse under the ADA's "regarded as" prong. According to the EEOC Manual, "'individuals who are subjected to discrimination on the basis of genetic information relating to illness, disease, or other disorders' fall under the third prong of the ADA's definition of disability."[286] A person may prevail under the "regarded as" prong if the employer mistakenly believes the individual has a disability when, in fact, the person does not, or if the employer mistakenly believes that an actual but non-limiting impairment substantially limits the individual in one or more major life activities.[287] The "regarded as" prong reflects Congress's belief that the reactions and stereotypes surrounding a disability are sometimes as harmful as the limitations that an actual impairment causes.[288]

To prove a "regarded as" claim, the claimant may rely upon, *inter alia*, the employer's statements and awareness of the employee's past medical history, evidence of employer prejudice, the employer's concerns about third-party prejudice or the employee's impact on the insurance program, stereotypes about non-disabling conditions, and the employer's failure to distinguish between a disability and lack of qualifications.[289] Decisions like

---

285. *Id.* at 653.

286. Deborah Gridley, Note, *Genetic Testing under the ADA: A Case for Employment Discrimination*, 89 GEO. L.J. 973, 989–90 (2001) (quoting 2 EEOC Compl. Man. (BNA) No. 198, at 902.0045 (Mar. 1995)).

287. Pagnatarro, *supra* note 199, at 161.

288. *Id.* at 161–62.

289. *Id.* at 163 (quoting Brian East, *The Definition of Disability after* Sutton v. United

Jones v. Inter-County Imaging Centers[290] suggest that an employer's concern regarding an employee or applicant's genetic code adversely impacting future attendance, productivity, or health insurance costs would have violated the ADA's "regarded as" prong prior to GINA.[291]

However, in order for a claimant to receive protection under the "regarded as" prong, the employer must act on the belief that the present or prospective employee has an impairment that presently and substantially limits a major life activity. In *Burlington Northern*, BNSF did not regard its employees as having a present limitation; it merely believed that the employees had a high risk of developing a disease in the future.[292] Therefore, the requirement for a present limitation would have posed a major bar to any victim alleging genetic discrimination under the ADA. Claimants would also have had to prove that the employer took an adverse employment action because it regarded the employee as a person with a disability.[293]

I. Direct Threat to Self Defense

Another reason that the pre-amended ADA inadequately protected against genetic discrimination in employment is that Section 12113(b) of the ADA provided employers with a "direct threat" defense that could easily preclude recovery for genetic discrimination.[294] A "direct threat" is an affirmative defense defined as a "significant risk to the health or safety of others that cannot be eliminated by reasonable accommodation."[295] In assessing the direct threat defense, courts consider the following factors: (i) "the duration of the risk," (ii) "the nature and severity of the potential

---

Airlines, Presentation Before the National AT Conference: Bridges to Better Advocacy (2000)).

290. 889 F. Supp. 741, 744 (S.D.N.Y. 1995) (determining that the plaintiff stated a valid cause of action under the ADA when he claimed that his employer fired him because the employer believed that the plaintiff's sickle cell anemia would adversely affect his future work attendance).

291. Miller, *supra* note 50, at 191; *see* Anderson v. Gus Mayer Boston Store of Del., 924 F. Supp. 763, 769 (E.D Tex. 1996) (employee with AIDS brought ADA claim regarding equal access to health insurance); Sawinksi v. Bill Currie Ford, Inc., 866 F. Supp. 1383 (M.D. Fla. 1994) (employee with injuries from brain tumor brought successful ADA claim after his termination).

292. *See supra* notes 215–23 and accompanying text.

293. Gridley, *supra* note 286, at 992; *see also* Cook v. R.I. Dep't of Mental Health, Retardation, and Hosps., 10 F.3d 17 (1st Cir. 1993) (regarding an employer who after refusing to hire a morbidly obese nurse successfully defended against her ADA "regarded as" claim on grounds that the employer did not regard her as substantially limited in the major life activity of working, but instead regarded her as unable to perform the particular job for which she applied, not a particular class of jobs or a broad range of jobs).

294. Silvestri, *supra* note 238, at 418.

295. *Id.* at 421 (internal quotation marks omitted).

harm," (iii) "the likelihood that the potential harm will occur," and (iv) "the imminence of the potential harm."[296] The EEOC states that the "qualification standard" includes both a self-threat and a "threat to others."[297] In the absence of GINA, a broad interpretation of this loophole might have allowed employers to discriminate against genetically predisposed individuals and co-dominant carriers by arguing that employment would exacerbate or trigger expression of their "bad" genes.

The United States Supreme Court broadly interpreted the "direct-threat-to-self" defense in *Chevron USA Inc. v. Echazabal*.[298] Echazabal worked in the oil refineries of an independent contractor for Chevron.[299] When he applied to work at Chevron directly, he received a job offer contingent on passing a physical exam.[300] After both Chevron's and Echazabal's doctors discovered that Echazabal had symptoms indicative of asymptomatic, chronic active Hepatitis C, Chevron withdrew its job offer.[301] None of Chevron's or Echazabal's doctors advised him to discontinue working in the coker unit of his current job.[302] In 1995, he applied again, and after failing the pre-entrance physical, Chevron told the independent contractor to "remove Mr. Echazabal from the refinery or place him in a position that eliminates his exposure to solvents/chemicals."[303] In response, Echazabal filed suit, alleging discrimination in violation of the ADA.[304]

The Supreme Court held that the "direct threat defense must be 'based on a reasonable medical judgment that relies on the most current medical knowledge and/or the best available objective evidence,' and upon an expressly 'individualized assessment of the individual's present ability to safely perform the essential functions of the job.'"[305] The Court found that although the ADA was intended to prevent paternalism by employers, an employer could refuse to hire a person because the person's condition or predisposition to developing a condition posed a direct threat to the employee, even if not to others.[306] The decision was especially surprising because Echazabal was asymptomatic and thus clearly able to perform the essential functions of his job without a reasonable accommodation and without imposing an undue hardship on the employer.[307]

---

296. *Id.*
297. *Id.* at 422 (internal quotation marks omitted).
298. 536 U.S. 73 (2002).
299. Silvestri, *supra* note 238, at 422.
300. *Id.*
301. *Id.* at 422–23.
302. *Id.* at 423.
303. *Id.*
304. *Id.*
305. *Id.*
306. *Id.* at 424.
307. *Id.*

Echazabal demonstrates why genetic discrimination could not have been adequately addressed by the ADA because employers can deny employment opportunities on the ground that worksite conditions or job duties will trigger development of diseases to which persons are genetically predisposed. "While ADA case law purports to require a 'high probability of substantial harm' before an employer can assert the direct threat defense, the Supreme Court's decision in *Echazabal* contravenes the wisdom of such precedent."[308] Echazabal's present condition did not substantially limit him, just as an employee genetically predisposed for breast cancer or carrying the gene for Huntington's disease is not presently limited by his or her genotype. "Instead of making the determination of the direct threat based on the current condition of the applicant or employee, *Echazabal* requires a consideration of the possible effect that the condition will have on that individual's 'future health.'"[309] While *Echazabal* did not directly deal with a genetic predisposition, its implications for genetic discrimination claims under the ADA are potentially far-reaching. Moreover, recent precedent suggests the courts' willingness to let employers off the hook via the direct threat-to-self defense.[310]

Significantly, it appears that no court has ever decided genetic discrimination in an employment case under the ADA.[311] However, the foregoing analysis makes clear that due to the exceptionally complex nature of genetics, GINA is better equipped to address problems unique to genetic discrimination in employment.[312] Moreover, other employment laws fail to negate the unique need for GINA.

---

308. *Id.* at 425.

309. *Id.*

310. *Id.* at 426–27 (discussing a disability case in which the court relied on *Echazabal* in stating that "the appellant's willingness to work is admirable, [but] we find that the consequences resulting from an accidental exposure could prove irreversibly catastrophic to her health"). In footnote 5 of *Orr v. Wal-Mart Stores, Inc.*, 297 F.3d 720 (8th Cir. 2002), the Eighth Circuit mentioned how the direct-threat-to-self defense would have changed the case:

> Had Orr established a prima facie case of actual disability under the ADA, Wal-Mart could have raised the threat-to-self defense. Wal-Mart could have argued that . . . working in a single-pharmacist pharmacy, which did not provide for uninterrupted meal breaks, posed a direct threat to Orr's health and that Wal-Mart was justified in not continuing his employment.

297 F.3d at 725. Orr was a diabetic. *Id.* at 722.

311. Miller, *supra* note 229, at 10.

312. For a different perspective, *see* Jennifer Chorpening, Comment, *Genetic Disability: A Modest Proposal to Modify the ADA to Protect against Some Forms of Genetic Discrimination*, 82 N.C. L. REV. 1441, 1475–80 (2004).

> [T]he Act should unambiguously allow employers to refuse to hire an applicant if the applicant's disease is such that if symptomatic the applicant would be barred

J. Title VII

Title VII forbids discrimination in employment on the basis of race, color, national origin, religion, and sex.[313] With respect to genetic discrimination claims, under disparate treatment and disparate impact theories, the employer can typically meet its burden by establishing that it had a legitimate business reason for testing and that the reason was not a pretext for discrimination.[314] To determine whether the employer's proffered reason is valid and justifiable, courts look to factors such as the nature of the business involved, the business practice at issue, and the degree of discriminatory impact.

If an employer's facially neutral genetic testing policy disproportionately affects a particular racial or ethnic group, such as Ashkenazi Jews,[315] persons of Mediterranean origin, or African-Americans,[316] the employer could arguably be sued under a Title VII disparate impact claim.[317] For instance, Ashkenazi Jews suffer higher rates of Tay-Sachs Disease, colon cancer, and possession of the BRCA-1 gene, which indicates a predisposition for breast cancer,[318] while African-

---

> from that particular job because of a direct threat to the applicant or others. . . . For those with genetic predispositions that would not normally come under the ADA when symptomatic, the employer would be allowed to choose whether to employ the individual without implicating the ADA. . . . Where an employee is predisposed to a disease that would not generally be considered a disability when symptomatic, such as a condition that can be corrected with medical treatments, the ADA would not apply.

Id. at 1476–77.

313. Title VII of the Civil Rights Act of 1964, 42 U.S.C. § 2000e-2(a) (2006).

314. 42 U.S.C. § 2000e-2(k).

315. Stolberg, *supra* note 90, at A24. Contrary to popular belief, Ashkenazi Jews do not have more defective DNA or generally higher rates of hereditary disease than other racial or ethnic groups. *Id.* Instead, "centuries of living and marrying within the confines of ghettoes [sic] have produced a relatively homogenous population in which tiny genetic alternations, or mutations, that cause disease are easy to find." *Id.*

316. Janet Brewer, *"Diseases of Place": Legal and Ethical Implications of Surname and Ethnicity as Predictors of Disease Risk*, 9 QUINNIPIAC HEALTH L.J. 155, 156–57 (2006). Sickle cell anemia illustrates the legal and ethical implications of surname, ethnicity, and race as predictors of disease risk. In the United States, one in twelve African-Americans carries the sickle cell gene, one in 400 babies diagnosed with sickle cell is African-American, and one in 1400 diagnosed with sickle cell is Hispanic. *Id.* at 157.

317. Kaufmann, *supra* note 239, at 418–21 ("[W]hen a disease occurs more frequently among a particular minority group . . . members of the minority group without the trait may be stigmatized and discriminated against solely because they are part of a group associated with the trait."); Miller, *supra* note 50, at 191.

318. Stolberg, *supra* note 90, at A24. Current estimates reveal that the genetic mutation known to cause breast cancer appears in 2.3% of Ashkenazi Jewish women, while 6% of Ashkenazi Jews carry the gene for colon cancer. *Id.* As of spring 1998, less than 1% of non-

Americans are far more likely to carry the sickle cell trait than other racial groups.[319] Similarly, women are disparately impacted by genetic testing for the BRCA-1 gene because breast cancer is likelier to occur in women, and such test results could lead to a disproportionate number of equally qualified female applicants being rejected arguably on the basis of a condition that, like pregnancy, could be a proxy for sex.

In *Griggs v. Duke Power Company*, the United States Supreme Court held that, under certain circumstances, facially neutral employment practices that disparately impact minority groups violate Title VII.[320] To establish a disparate impact claim, the employee must show that the genetic testing or screening disparately impacted a protected group.[321] The burden then shifts to the employer to prove that the testing or screening served a legitimate business purpose.[322] Once met, the burden of proof shifts back to the employee who must prove that an equally effective, viable alternative was available that would have had a smaller disparate impact on the protected group.[323]

However, even if a disparate impact theory was successful, Title VII would still have provided insufficient protection against genetic discrimination in employment because Title VII only protects members of a protected class, and many genetically related conditions and disorders do not disproportionately impact specific races, ethnic groups, or sexes.[324]

K. OSHA

In 1970, Congress enacted the Occupational Safety and Health Act ("OSHA") in order to ensure "safe and healthy" worksites.[325] OSHA's General Duty Clause requires employers to provide a workplace "free from hazards likely to cause death or serious physical harm."[326] Critics of GINA

Jews had been found to carry the mutation for breast cancer and the colon cancer gene had not been found in any non-Jews. *Id.* Ashkenazi Jews with the mutation have an 18 to 30% chance of developing colon cancer whereas the rate of colon cancer development for the general population is only 5%. *Id.* A woman with no breast cancer mutation has a 13% chance of developing breast cancer as compared to a woman with any of the three mutations linked to breast cancer who has a 56% chance of developing that condition. *Id.*

319. Brewer, *supra* note 316, at 156–57. As of April 1997, geneticists discovered that one could distinguish between blacks and whites on the basis of differences in just three out of 100,000 human genes. Andrews, *supra* note 13, at 51. *See generally* 2 ENCYCLOPEDIA OF BIOETHICS, *supra* note 27, at 992–96 (discussing the high rates of sickle cell in the African American community and possible historical reasons for the strong genetic correlation).

320. 401 U.S. 424, 436 (1971).

321. Kaufmann, *supra* note 239, at 421.

322. *Id.*

323. *Id.* at 421, 423–24.

324. Miller, *supra* note 50, at 191.

325. Taylor, *supra* note 50, at 57–58.

326. *Id.* at 58.

argued that OSHA regulations concerning employee medical exams already governed the use of employment-related genetic tests.[327] Employers claimed that OSHA not only permits but also requires them to use genetic screening and monitoring to ensure that genetically predisposed employees are not given job assignments or worksite locations that will aggravate their genetic disorders or trigger the manifestation of a disease for which the individual is genetically predisposed.[328]

Yet OSHA, standing alone, inadequately protected against genetic discrimination in employment for several reasons. First, OSHA regulates "medical exams," but genetic tests given to employees for purposes unrelated to medical treatment or diagnosis might not fall under this classification.[329] Second, OSHA's medical exams are "hazard-specific" and are only required when employees are exposed to certain substances like lead or cotton dust.[330] Most importantly, prior to GINA's enactment, OSHA announced that "its standards do *not* require genetic testing in the workplace," only that the employee's health status be "identifiable," which can be accomplished by other less intrusive means such as physical exams.[331] Consequently, OSHA provides minimal guidance on genetic testing and little, if any, protection against genetic discrimination in employment.[332]

That being said, GINA addresses employer concerns regarding employee monitoring by permitting employers to request or require genetic information to monitor the biological effects of toxic substances in the workplace but forbids employers from basing employment decisions on that information.[333] GINA protects the privacy of employees by allowing the employer to receive monitoring results only in an aggregate form, by requiring employers to give written notice of monitoring to the covered individual, and if monitoring is not required by law, to obtain the employee's prior, knowing, voluntary, written authorization.[334] GINA requires employers to inform covered individuals of the monitoring results and to ensure that the monitoring complies with applicable laws and regulations.[335]

---

327. *Id.*
328. *Id.*
329. *Id.*
330. *Id.*
331. *Id.* at 58–59 (emphasis added).
332. *Id.* at 59.
333. *See supra* notes 172–75 and accompanying text.
334. *See supra* notes 172–75 and accompanying text.
335. *See supra* notes 172–75 and accompanying text.

L. Executive Order No. 13145

On February 8, 2002, former President William J. Clinton issued Executive Order No. 13145, which forbade the use of genetic testing as a condition of employment or promotion for prospective and present employees of the executive branch or its agencies.[336] Only Congress can bequeath like protection on the private sector, so the Executive Order only applies to federal employees.[337] While the Executive Order sent a powerful message to Congress regarding the President's stance on genetic discrimination in employment, it failed to adequately protect the public against genetic discrimination in employment because it covered only a narrow subset of American workers.[338]

M. State Legislation

Some critics of GINA claim that genetic discrimination is an issue best left to state and local regulation.[339] However, as of 2006, only thirty-two states had enacted laws prohibiting genetic discrimination in employment.[340] Such laws typically take two different approaches:

---

336. Ellis, *supra* note 201, at 1087.

337. *Id.* at 1088.

338. *Id.*

339. *See supra* note 224 and accompanying text.

340. Slaughter, *supra* note 10, at 73. For a chronological perspective on the enactment of state genetic discrimination laws, *see* Andrews, *supra* note 13, at 45 (As of April 1997, only six states had laws that prohibited genetic testing in employment without consent.). For a discussion of Massachusetts's passage of a genetic discrimination law, *see* Tyson, *supra* note 193, at 3. For an excellent timeline of state and federal legislation that regulates genetic information, *see* Calvo, *supra* note 193, at 30. As of September 2000, twenty state laws restricted the disclosure of genetic information to third parties; thirteen state laws limited acquisition, retention, and disclosure to third parties of genetic information; and five states defined genetic information as personal property. *Id. See generally* CAL. INS. CODE § 10123.3 (2006) (prohibiting discrimination on the basis of genetic characteristics); 19 DEL. CODE ANN. tit. 19, § 711 (2006) (prohibiting employment discrimination on the basis of race, marital status, genetic information, color, age, religion, sex or national origin); IOWA CODE ANN. § 729.6 (West 2005) (regulating "genetic testing"); LA. REV. STAT. ANN. § 23:368 (2007) (prohibiting genetic discrimination in the workplace); MASS. ANN. LAWS ch. 175, § 108H (Lexis Nexis 2006) (defining "genetic information" and "genetic testing"); MASS. ANN. LAWS ch. 175, § 120E (Lexis Nexis 2006) (prohibiting "genetic based" discrimination); MASS. ANN. LAWS ch. 175, § 1081 (Lexis Nexis 2006) (describing discrimination on the basis of genetic information or a genetic test as "unfair discrimination"); ME. REV. STAT. ANN. tit. 5 § 19302 (2006) (prohibiting employment discrimination on the basis of genetic information or genetic testing); ME. REV. STAT. ANN. tit. 24-A, § 2159C (2006) (prohibiting discrimination on the basis of genetic information or testing); N.J. STAT. ANN. § 10.5-44 (2007) (prohibiting the collection, retention, or disclosure of DNA without consent); N.Y. INS. LAW § 2615 (McKinney 2006) (requiring informed consent in order to obtain genetic testing); N.C. GEN. STAT. § 95-28.1A (2006)

exceptional and inclusive.[341] The exceptional approach addresses genetic information in separate legislation, whereas the inclusive approach treats genetic information like other protected traits, such as race or sex.[342]

These non-uniform state laws were insufficient and created inconsistency, confusion, and inequity, making it extremely difficult for multi-state employers to understand and comply with state laws.[343] To

---

(prohibiting discrimination against person based on genetic testing or genetic information); OR. REV. STAT. § 659A.303 (2006) (prohibiting employers from obtaining, seeking to obtain, or using genetic information); TEX. LAB. CODE ANN. § 21.402 (West 2006) (prohibiting "discriminatory use of genetic information"); VA. CODE ANN. § 38.2-508.4 (2006) (regulating "genetic information privacy").

341. Silvestri, *supra* note 238, at 419. For a survey of different approaches to statutory prohibitions against genetic discrimination in employment, *see* Feldman & Katz, *supra* note 128, at 411–16.

342. Silvestri, *supra* note 238, at 419–20. Maryland illustrates the exceptional approach by addressing genetic discrimination in a separate law that forbids employers from requiring employees to undergo genetic testing as a condition of employment. *Id.* Maryland also forbids the use of genetic information in calculating wages, bonuses, and raises and specifically prohibits employment discrimination on the basis of a genetic predisposition. *Id.* at 420–21.

343. Ellis, *supra* note 201, at 1073; *see* Tyson, *supra* note 193, at 3; *see also* Chandrani Ghosh, *Employers Beware*, FORBES, Oct. 2, 2000, at 56 (characterizing state laws prohibiting genetic discrimination in employment as "punishing" the private sector and "meddling in the workplace"). Statutory approaches vary:

> Arizona prohibits discrimination only if it is "based on the results of a genetic test received by the employer." Connecticut bars employers from requesting or requiring genetic information from either employees or individuals seeking employment. In Delaware, it is an unlawful employment practice for employers to consider genetic information in making employment decisions . . . [and] "limit[ing], segregat[ing], or classify[ing] employees" in a manner that may deprive them of "employment opportunities or otherwise adversely affect . . . [their] status." . . . [I]n Florida . . . statutory protection is extended to prospective employees denied employment predicated on genetic information. . . . Iowa prohibits employers from asking, demanding, or giving a genetic test to an individual as a condition precedent to employment. Similarly, employers are prohibited from changing "the terms, conditions, or privileges of employment" of anyone who undergoes a genetic test. Iowa's law does not have plain language prohibiting employers from using genetic information in their employment decisions. . . . Maine prohibits employment discrimination on the basis of genetic information or genetic testing . . . [as well as] employees who refuse to submit to a genetic test or make tests results available . . . Michigan [employers] may not discriminate against individuals because of "genetic information that is unrelated to the individual's ability to perform the duties of a particular job or position." Missouri allows employers to "distinguish between [or] discriminate against" an individual based on genetic information when the information is "directly related to his or her ability to perform a job."

complicate matters, many state laws are woefully under-inclusive in that they apply only to carriers of certain gene mutations, such as sickle cell anemia, or only prevent discrimination due to "genetic testing," not "on the basis of genetic information derived from any source."[344] For example, Rhode Island prohibited employers from requiring or requesting genetic information from a prospective or present employee.[345] Other states allowed employers to request and require genetic information.[346]

To further complicate matters, state laws defined genetic information and genetic testing quite differently.[347] Statutes that narrowly define genetic information are easier to implement but often provide insufficient protection because employers may still discriminate based on substituted indirect tests or family history.[348] On the other hand, laws with broader definitions that include gene products and family history are often more difficult to implement. For instance, George Annas has argued that a broader definition may even require an "overhaul of well-established medical information practices and policies."[349] Accordingly, GINA was necessary to establish a uniform standard that would "fully protect the public from discrimination and allay their concerns about the potential for discrimination, thereby allowing individuals to take advantage of genetic testing, technologies, research, and new therapies."[350]

N. Genetic Exceptionalism

Although GINA is not impervious to criticism, the relationship of genetic information to other protected traits also justifies GINA's enactment. The exceptional nature of genetic information and the unique problems inherent in its regulation necessitate customized legislation.

Feldman & Katz, *supra* note 128, at 412–13 (citations omitted).

344. Shoop, *supra* note 19, at 14.

345. Ellis, *supra* note 201, at 1088; *see* Michael R. Santiago, *Review of Selected 1998 California Legislation—The Industry of Death: Regulating Mortuary Services*, 30 MCGEORGE L. REV. 463 (1999).

346. Ellis, *supra* note 201, at 1088.

347. Taylor, *supra* note 50, at 67 (discussing New York law that prohibits discrimination on the basis of genetic predisposition or carrier status); *see* Feldman & Katz, *supra* note 128, at 411. Two states exclude medical history from the definition of genetic information, and several other states exclude blood and urine tests. *See* Michael S. Yesley, *Protecting Genetic Difference*, 13 BERKELEY TECH. L.J. 653, 659–61 (1998). Roughly half of the definitions include tests for gene products in the term "genetic testing." *See id.* Some states use different definitions in the employment versus insurance contexts. *See id.*

348. Yesley, *supra* note 347, at 659–61.

349. *Id.* at 661.

350. Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-233 § 2(5), 122 Stat. 881, 882–83 (2008) (codified as amended at 42 U.S.C. §§ 2000ff to 2000ff-12 (Supp. III 2009)).

Genetic information has much in common with other protected traits like race, sex, and religion. For instance, just as each of us has "bad" genes, which make us susceptible to genetic discrimination, each of us also has a race, national origin, and sex, which could do the same. Just as certain types of discrimination affect portions of the population disproportionately, so too, does genetic discrimination have a disparate impact on certain groups.[351]

In 1971, geneticist Bentley Glass announced "the right of every child to be born with a sound physical and mental constitution, based on a sound genotype," and stated that in the future, parents will have no "right to burden society with a malformed or a mentally incompetent child."[352] As Glass's remarks illustrate, genetic discrimination stigmatizes its victims, creating a "genetic underclass" in some ways as psychologically damaging as the "racial underclass" that inspired the Civil Rights Movement. Social psychologist Claude Steele studied "stereotype vulnerability"—a phenomenon in which students perform more poorly if informed *ex ante* that they are members of a social group that in the past has not been academically strong.[353] Lori Andrews implies that informing employees that they belong to a genetic underclass might cause "stereotype vulnerability" among affected portions of the workforce, causing members of that underclass to perform more poorly than they would have before learning of their genetic anomalies.[354] Stereotype vulnerability would not only act as a self-fulfilling prophecy preventing the "healthy ill" from realizing their full work potential but would also adversely impact employers via lower morale and decreased productivity and output among workers.[355]

Like racism and sexism, genetic discrimination is not solely a twenty-first century phenomenon. Instead, it is rooted in a long, dark history of eugenics—a movement that perhaps began with Sir Francis Galton in which scientists and social reformers launched a campaign of coercive sterilizations against the poor, infirm, incarcerated, and mentally ill to prevent and discourage what they perceived to be the genetically defective underclass from reproducing (*i.e.*, negative eugenics), while simultaneously encouraging procreation by those thought to have superior genetic endowments (*i.e.*, positive eugenics).[356] In 1907, Indiana passed America's

---

351. *See supra* notes 93, 315, 318 and accompanying text.

352. Kass, *supra* note 95, at 40.

353. Andrews, *supra* note 13, at 45; *see* Weaver, *supra* note 233, at 243.

354. Andrews, *supra* note 13, at 48.

355. *Id.*

356. *See* Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-233 § 2(2), 122 Stat. 881, 882 (2008) (codified as amended at 42 U.S.C. §§ 2000ff to 2000ff-12 (Supp. III 2009)) (discussing eugenics); 2 ENCYCLOPEDIA OF BIOETHICS, *supra* note 27, at 1021.; *see also* Edward J. Larson, *The Meaning of Human Gene Testing for Disability Rights*, 70 U. CIN. L. REV. 913 (2002) (discussing the stigmatization of the disabled

first sterilization law, which targeted persons suffering from several genetic disorders, and in 1910, the Eugenics Record Office was established to collect genetic information for the purpose of advising Americans on who was "fit" for marriage.[357] In 1927, the United States Supreme Court upheld Virginia's sterilization law against a substantive due process challenge in the infamous case, *Buck v. Bell*, which has never been overruled.[358]

Eugenics had racial, ethnic, and classist overtones. The "better" genetic stock consisted of white, middle, and upper class Protestants of European descent, while African-Americans, Native Americans, Jews, Catholics, and Southern European immigrants were considered to be "inferior" stock.[359] One eugenicist explained the movement's logic as follows: "The superficially sympathetic man flings a coin to the beggar; the more deeply sympathetic man builds an almshouse for him so that he need no longer beg; but perhaps the most radically sympathetic of all is the man who arranges that the beggar not be born."[360] Only after World War II did American scientists and the public disavow eugenics, largely due to its association with the Nazi atrocities that purported to be "genetic hygiene" experiments aimed at exterminating the Jews in order to create a pure Aryan race.[361]

For these reasons, "geneticism" (*i.e.*, using one's genetic information to subordinate or discriminate against him or her) is arguably just as harmful to societal interests as its better known counterparts—racism and sexism—and may send a message to the "genetic underclass" that they would have been better off having never been born because their genetic code renders them incapable of making any valuable contributions to society.[362]

---

throughout U.S. history, including the early eugenics movements, which attempted to sterilize mentally and physically disabled persons). The most illustrative example of the early eugenics movement was a eugenics-based film entitled "Are You Fit to Marry?" ARE YOU FIT TO MARRY? (Quality Amusement Corp. 1927). The movie, which was inspired by a physician-eugenicist who purportedly facilitated the death of newborns with genetic defects, centers on a wealthy man who carries a "bad" but unexpressed gene. *Id.* He marries without telling his new wife of his genetic status, and the couple soon has a child with a disability. *Id.* The doctor tries to convince the parents to allow the child to die, but they refuse. *Id.* The story follows the child's life, showing him taunted by others and suffering from his medical condition. *Id.* The movie ends with the child growing up to become a derelict on the streets. *Id.*

357. Kaufmann, *supra* note 239, at 401–02.

358. Buck v. Bell, 274 U.S. 200 (1927) (Justice Holmes remarked, "Three generations of imbeciles is enough!").

359. Deborah Hellman, *What Makes Genetic Discrimination Exceptional?*, 29 AM. J.L. & MED. 77, 107 (2003); *see* Bromfield, *supra* note 53, at 126–30.

360. Hellman, *supra* note 359, at 107 (quoting DANIEL J. KELVES, IN THE NAME OF EUGENICS: GENETICS AND THE USES OF HUMAN HEREDITY 90 (1985)).

361. *Id.*

362. *Id.* at 90–91.

Although the foregoing sections have provided support for GINA's enactment, employers and employees have competing interests with regard to the regulation of genetic information. GINA's drafters had the arduous task of creating a statute that harmonized these interests, but did they succeed?

IV. AMERICA'S BRAVE, NEW WORKPLACE: COMPETING CONCERNS AND INTERESTS

A. Employer Concerns and Interests

As Maryland Delegate Dan Morhaim quipped, "[g]enetic technology is the life science equivalent of nuclear power,"[363] and its impact on employment will likely be just as explosive. A 1989 Office of Technology Assessment survey stated that 5% of American companies admitted to genetically monitoring and screening employees.[364]

As Director of the Health, Law, and Policy Institute, Mark Rothstein explained that "[e]mployers have a tremendous economic incentive to discriminate based on perceived future health status. They've done it with other conditions and we have every reason to expect they would do it in the genetics area."[365] There are numerous reasons why employers may desire access to their employees' genetic information: (i) to protect employee safety and health; (ii) to save money by avoiding high risk employees;[366] and (iii) to protect themselves against future liability for the possible adverse health consequences that exposure to hazardous workplace chemicals might have on employees genetically predisposed for certain health conditions.[367] At least some courts have held that avoidance of potential future liability amounts to a business necessity and thus, is a defense under the ADA.[368]

---

363. Calvo, *supra* note 193, at 28.

364. Andrews, *supra* note 13, at 47.

365. Payne, *supra* note 211, at 7.

366. *See Genetic Testing Raises Fears of Workplace Bias*, *supra* note 20, at 8H. As Law Professor Mark Rothstein explained, 5% of claimants represent 50% of insurance costs, so "[i]f an employer could eliminate big health-care users, it would be extremely beneficial." *Id.*

367. Elaine Draper, *The Screening of America: The Social and Legal Framework of Employers' Use of Genetic Information*, 20 BERKELEY J. EMP. & LAB. L. 286, 289 (1999); *see also* 2 ENCYCLOPEDIA OF BIOETHICS, *supra* note 27, at 950 (discussing how some states impose a duty to warn foreseeable third parties of imminent physical harm arising from the patient's genetic condition on genetic counselors and physicians while other states suggest that employers who are aware of an employee's genetic condition may be permitted or even compelled to disclose that information over the employee's consent if doing so is required to protect third parties from harm).

368. Brian M. Holt, Comment, *Genetically Defective: The Judicial Interpretation of the Americans with Disabilities Act Fails To Protect Against Genetic Discrimination in the*

Perhaps most importantly, employers want to ensure that employees are fit to perform the essential functions of their jobs. "If a worker will become ill, and if the employer will be responsible for the medical costs as well as the output costs of the worker's absence, then the predicted illness is nothing but a future dollar cost that the employer must consider and discount."[369] Unhealthy employees pose huge costs to employers in the form of above-average absenteeism, decreased productivity, overtime payments to hire workers to cover absent employees' shifts, higher job turnover, administrative costs inherent in hiring, recruiting, and training replacements, and higher workers' compensation insurance premiums that result when an employee makes a claim for benefits.

Notably, in 1992, 68% of employees in the United States had insurance through their employers.[370] Under self-insurance plans, which many large employers use, the employer is directly responsible for employees' health care expenses; therefore, as health care costs increase, employers who are not allowed to screen employees genetically predisposed to developing cancer, diabetes, or occupational illnesses may choose to eliminate health insurance altogether unless required by law to provide it.[371] Because large and medium-size employers are "experience-rated," their premiums increase as more employees make workers' compensation claims against them. Pooling employees' genetic information could protect each employee's privacy by ensuring anonymity.[372] Simultaneously, it would allow the employer's insurer to better assess the health and health risks of its workforce, more accurately adjusting the employer's insurance rate via experience-rating and providing it with a more accurately-priced policy.[373] Consequently, employers can better assess how much of that cost to pass on to consumers, perhaps resulting in lower product prices.

Title I of GINA addresses this concern by prohibiting genetic discrimination in insurance. Although a full discussion of Title I exceeds the scope of this Article, Title I prohibits a health insurance issuer offering health insurance coverage in the individual market from establishing eligibility rules of enrollment based on genetic information or from imposing any pre-existing condition exclusion on the basis of genetic

---

Workplace, 35 J. MARSHALL L. REV. 457, 473 (2002).

369. Larry Gostin, *Genetic Discrimination: The Use of Genetically Based Diagnostic and Prognostic Tests by Employers and Insurers*, 17 AM. J.L. & MED. 109, 133 (1991) (quoting Liebman, *Too Much Information: Predictions of Employee Disease and the Fringe Benefit System*, 1988 U. CHI. LEGAL F. 57, 82).

370. Steinforth, *supra* note 261, at 985.

371. *See id.*

372. Kathleen C. Engel, *Can Employers Put Genetic Information to Good Use?*, 16 J.L. & HEALTH 9, 13 (2002).

373. *Id.* at 13. *See generally* Draper, *supra* note 367 (analogizing genetic screening to existing drug screening policies and arguing that both screening measures identify problem employees while raising the employees' privacy concerns).

information.[374] It prohibits a group health plan from adjusting premium or contribution amounts for a group on the basis of genetic information and provides that the prohibition does not limit a group health plan's health insurance carrier from adjusting the group policy premium based on the manifested disease of a covered person.[375] However, the insurer may not use the manifested disease to increase the employer's premium.

Although a group health plan may not request or require a person to undergo genetic testing, a health care professional may request a person to submit to genetic testing.[376] The prohibition also does not forbid a group health plan from obtaining or using the results of a genetic test to make a determination regarding payment, so long as the plan only requests the minimum amount of information necessary.[377] Furthermore, a group health plan may not request, require, or purchase genetic information for underwriting purposes or with respect to any individual prior to such individual's enrollment in connection with the enrollment.[378]

Permitting employers to use genetic information and to conduct genetic testing and monitoring could also improve workplace safety. First, testing could better ensure employer compliance with OSHA, which requires that all worksites "be sufficiently 'free from recognized hazards that are causing or are likely to cause death or serious physical harm.'"[379] Employers comply by enacting safety measures that are "necessary or appropriate to provide safe or healthful employment."[380] Reliance on *ex post* monitoring can be very expensive because in many cases the quantity and quality of an employee's output cannot be observed or verified.[381] Taking remedial action also has indirect costs such as the harm to employee morale or the threat of litigation under Title VII or under a wrongful discharge tort claim.[382] *Ex ante* screening is preferable because *ex post* monitoring and corrective actions require investments in increased supervision of suspected

374. Summary of Genetic Information Nondiscrimination Act of 2008 (GINA), http://www.dnapolicy.org/resources/GINATitle1summary.pdf (last visited Aug. 30, 2011).

375. *Id.*

376. *Id.*

377. *Id.*

378. *Id.*

379. Steinforth, *supra* note 261, at 974 (quoting Occupational Safety and Health Act, 29 U.S.C. § 654(a)(1) (2001)).

380. *Id.* at 975 (quoting Indus. Union Dep't v. Am. Petroleum Inst., 448 U.S. 607, 615 (1980)).

381. Diver & Cohen, *supra* note 196, at 1461.

382. Many employers fear taking *any* adverse employment actions against an employee whose race, sex, religion, national origin, or color might give rise to a Title VII claim, even if frivolous. Hiring counsel to dismiss a frivolous claim on a 12(b)(6) motion is time-consuming, expensive, and can harm the employer's reputation as well as damage employee morale.

employees while the underperformance may be damaging morale and incurring opportunity costs in foregone output.[383]

Information derived from genetic testing and monitoring can also improve worker productivity and, consequently, the efficiency of labor markets by improving the match between each worker's qualifications and particular job requirements and facilitating more cost-effective investments in human capital.[384] The information collected from employment-related physical exams, family histories, and questionnaires could someday be entirely replaced by a single set of genetic tests that would arguably provide more accurate and reliable information.[385] Armed with this information, the employer could better predict each worker's intensity and quality of effort.[386]

While some argue that the best predictor of future performance is past performance, this argument fails to adequately take into account transaction costs.[387] Employers consider the employee or applicant's propensity for requiring future sick leave, treatment, early retirement, family and medical leave, reassignment, absence, and removal because those actions generate huge transaction costs for the employer, even if the employee or applicant is presently asymptomatic.[388] For example, hiring or promoting an under-qualified or under-productive worker to replace the employee whose genetic condition requires him to take sick leave or go on disability has a huge cost for the employer.[389] Advertising to recruit a replacement, expending time and energy to interview and evaluate applicants, and training a new hire are all employer expenditures that the employer could easily avoid if the employer knows *ex ante* that Employee A—a 39-year-old man who has the gene for Huntington's Disease—is very likely to develop symptoms, such as decreased cognition and adverse changes in temperament, in the near future and long before the employer will have received any return on his investment in interviewing, hiring, and training Employee A.[390]

Although the aforementioned use of genetic information is exactly what GINA aims to prevent, allowing employers to use genetic information, testing, and monitoring in other ways could actually benefit employees. Genetic monitoring could allow earlier detection of the impact of exposure to hazardous chemicals or radioactive substances on employees, perhaps

383. Diver & Cohen, *supra* note 196, at 1461.

384. *Id.* at 1460. *See generally* Engel, *supra* note 372 (arguing that in some situations, employers can use employees' genetic information for socially valuable purposes that benefit both employers and employees).

385. Diver & Cohen, *supra* note 196, at 1461.

386. *Id.*

387. *Id.*

388. *Id.* at 1461.

389. *Id.*

390. *Id.*

allowing employers to enact increased health and safety precautions or remove at-risk employees to low-exposure work areas.[391] Alternately, employers could more accurately determine whether a worker's under-performance was due to a genetic or medical factor or other causes, such as laziness or carelessness. Then employers could remediate the problem via reassignment rather than undeserved disciplinary actions like suspension, demotion, or termination. Employers could also use genetic information to measure deficits in job-related skills and correct them through training programs targeted to meet the specific needs of individual employees.[392]

Though some argue that allowing employers to reassign pre-disposed workers to safer work areas diminishes employer incentives to remove workplace hazards, that argument fails because other incentives to make the workplace safer would remain, including OSHA compliance and experience-rating.[393] Likewise, employers want to avoid the reputational harms generated by occupational accidents and diseases because such harms require employers to expend more resources in recruiting and retraining replacements as well as retaining current employees. A reputation as "dangerous" will also damage morale and thus, hinder productivity. Clearly then, even if some genetic testing were permissible, incentives to improve workplace safety would remain.

GINA addresses some of these concerns by allowing covered entities to request or require information to genetically monitor the biological effects of toxic workplace substances so long as the following conditions are met: (i) the covered entity gives written notice to covered individuals; (ii) the monitoring complies with applicable federal, state, and local laws; (iii) the employer receives only aggregate monitoring results that do not identify individual participants; and (iv) where the monitoring is not required by law, the employer obtains the individual's prior, knowing, voluntary, and written authorization.[394] Genetic information garnered from genetic monitoring may not be used to make employment decisions or to limit, classify, or segregate employees in a manner that will deprive or tend to deprive them of employment opportunities, and as an extra measure to protect employee privacy, such information must be stored in a separate medical file marked confidential.[395]

Allowing employers to use genetic information, testing, and monitoring in other ways could actually benefit employees. Such testing might accurately assess the risk that a person may pose to the workplace and vice versa.[396] After identifying that certain portions of its workforce are predisposed to certain health risks, concerned employers could engage in

391. Ellis, *supra* note 201, at 1090.
392. Diver & Cohen, *supra* note 196, at 1462.
393. *Id.* at 1462–63.
394. 42 U.S.C. § 2000ff-1(b)(5) (Supp. III 2009).
395. *Id.* § 2000ff-1(a).
396. Ellis, *supra* note 201, at 1090.

"self-interested philanthropy," designing free (either mandatory or optional) fitness or wellness programs aimed at reducing behaviors that aggravate those health risks.[397] For instance, employers could install fitness centers at the worksite and allow employees an exercise break, just as they currently permit unpaid lunch and rest breaks. Employers could hire nutritionists to design company cafeteria entrees that either prevent heart disease or at least, alleviate the symptoms of the disease. They could monitor indicators of employee health at regular intervals and even redesign production and the worksite so as to prevent exposure to environmental triggers. Employers would incur costs for these new programs but in return, would experience improved morale, increased productivity, reputational benefits, lower administrative costs, decreased use of employee sick leave, greater returns on investments in recruiting and training employees, reduced absenteeism, and long-term reductions in insurance payouts. A happier and healthier workforce would simultaneously benefit employers and employees.

GINA addresses this issue by permitting a covered entity to provide health or genetic services to a covered individual (including a participant in a wellness program) so long as the person provides prior knowing, voluntary, and written authorization. Only the participant and the healthcare professional involved receive individually identifiable information concerning the results of the genetic services.[398] GINA protects participants' genetic privacy by forbidding the disclosure of identifiable information, except in an aggregate form that will not identify the participant; furthermore, no adverse use may be made with the information.[399] Under GINA, an employer-sponsored voluntary wellness program may not require genetic information or penalize individuals who refuse to provide it.[400] Nor may a covered entity offer a financial inducement for the provision of genetic information.[401] Such programs must be offered to persons with existing health conditions and those whose lifestyle choices increase their risk of developing a condition.[402]

Genetic testing may also determine whether an employee's health condition poses a direct threat. Prior to GINA, under the reasoning of *Echazabal*,[403] an employer could legally fire or refuse to hire a person whose genetic predisposition or condition would be exacerbated or triggered by exposure to certain workplace chemicals. Arguably, such

397. Diver & Cohen, *supra* note 196, at 1462.

398. 42 U.S.C. § 2000ff-1(b).

399. *Id.*

400. Genetic Information Nondiscrimination Act of 2008, 29 C.F.R. § 1635.8 (2011).

401. *Id.* However, it is permissible to offer financial inducements for the completion of health risk assessments that include questions about family medical history or other genetic information, so long as the covered entity clarifies that the inducement is available whether or not the participant answers questions regarding genetic information.

402. *Id.*

403. *See supra* notes 299–310 and accompanying text.

testing is job-related and consistent with business necessity. If an employer knew that Employee A was genetically predisposed to heart disease and Employee A exhibits chest pains at work, his employer could alert paramedics to that information, perhaps saving Employee A's life or at least giving his treating physicians a clue as to what may be the source of Employee A's symptoms.[404] Neither GINA nor its accompanying regulations explicitly address this issue.

B. Employee Concerns and Interests

As explained above, prior to GINA and perhaps still today, many employees feared that their employers would use their genetic information to discriminate against them, so they opposed allowing employer access to their genetic information.[405] One reason that employees object to genetic testing in employment is that genetic testing is not 100% accurate.[406] Assuming the test is reliable and properly administered, test results are of limited significance because a person who carries genetic markers predisposing her to some disorder may or may not develop it.[407] Even if the disorder's heritability is 50%, the person may never encounter the environmental stimulus required to trigger phenotypic expression of the disorder or may die before the disorder's age of onset.[408]

Prior to GINA's enactment, genetic screening and monitoring of employees was sometimes "voluntary" in name only. For example, in *Burlington Northern & Santa Fe Railway*, an employee who refused testing was allegedly threatened with imminent discharge unless he consented.[409] Likewise, some applicants were purportedly pressured to undergo genetic tests as part of their pre-entrance employment exams, while employees were threatened with loss of insurance, termination, or demotion unless they consented.[410] Independent genetic counseling is often unavailable, and without an adequate understanding of genetics, perhaps consent is not truly informed.[411]

As mentioned above, GINA protects against this possibility by allowing covered entities to request or require information to genetically monitor the biological effects of toxic substances in the workplace only if they give written notice of the monitoring to covered individuals, and, if the monitoring is not required by law, obtain the person's prior, knowing,

---

404. Engel, *supra* note 372, at 12.
405. *See* 2 ENCYCLOPEDIA OF BIOETHICS, *supra* note 27, at 948–52.
406. *See supra* notes 136–41 and accompanying text.
407. *See supra* notes 136–41 and accompanying text.
408. *See supra* notes 136–41 and accompanying text.
409. Miller, *supra* note 50, at 10.
410. Draper, *supra* note 367, at 293.
411. *Id.*

voluntary, written and authorization.[412] The covered entity must inform the covered individual of the results of the monitoring and ensure that the monitoring complies with applicable laws and regulations.[413] The covered entity may only receive the monitoring results in an aggregate form that does not identify the monitored individual.[414]

Additionally, in the absence of GINA and perhaps even after its enactment, genetic technology may have led to "medicalization," in which "what once was regarded as a normal behavior or bodily state now is regarded as abnormal because there are medical interventions that give people control over that behavior or state."[415] As Leon Kass elucidates:

> [T]he standard of health is being deconstructed. Are you healthy if, although you show no symptoms, you carry genes that will definitely produce Huntington's disease? What if you carry, say, 40 percent of the genetic markers thought to be linked to the appearance of Alzheimer's disease? And what will "healthy" and "normal" mean when we discover your genetic propensities for alcoholism, drug abuse, pederasty, or violence? The idea of health progressively becomes at once both imperial and vague: medicalization of what have hitherto been mental or moral matters paradoxically brings with it the disappearance of any clear standard of health itself.[416]

GINA may fail to prevent "medicalization" because research reveals that despite the existence of the ADA, applicants with disabilities on average receive fewer interview callbacks, less favorable hiring recommendations, lower salary recommendations, and lower ratings than non-disabled applicants.[417]

Moreover, mandatory genetic screening and monitoring policies that perhaps could have existed in the absence of GINA entirely overlook an employee's right *not* to know his genetic makeup.[418] For instance, a 1996

412. 42 U.S.C. § 2000ff-1.

413. *Id.*

414. *Id.*

415. 2 ENCYCLOPEDIA OF BIOETHICS, *supra* note 27, at 982.

416. Kass, *supra* note 95, at 44.

417. Adya & Bornstein, *supra* note 208, at 283–85 (noting that employers offer amputees and epileptics lower salaries than similarly qualified non-disabled persons and describing disabilities as "stigmas in organizations"). *See* Larson, *supra* note 356 (discussing the stigmatization of the disabled throughout United States history, including the early eugenics movements); Kass, *supra* note 95 (discussing an instance in which a doctor told a group of medical students that had a ten-year-old boy with spina bifida been conceived today, he would have been aborted).

418. Weaver, *supra* note 233, at 243–44 (discussing a man whose father was diagnosed with Alzheimer's disease and admitted, "I do not want to be screened for the Alzheimer's gene. I am an 'avoider.' The potential discrimination that I would face far outweighs any benefit this knowledge might give me."); *see also* Katherine A. Schneider, *Adverse Impact of Predisposition Testing on Major Life Activities from BRCA1/2 Testing*, 3 J. HEALTH CARE

study reported that roughly 50% of participants with a family history of breast or ovarian cancer stated that they would not want to know if they possessed a genetic predisposition for those diseases.[419] Learning one's genetic code can elicit a wide range of emotions from elation to guilt because the individual is free from a genetic curse but one's beloved brother or sister, who shares half of the same genes, is not.[420] The hopelessness of learning that you are one of the "healthy ill" could lead to psychological effects, such as depression and anxiety.[421] Knowing one's genetic fate may also negatively impact one's aspirations and desire to marry, have children, attend graduate school, or pursue other long-term goals.[422] Because many genetic tests detect disorders for which no treatment exists, a tested individual may make major lifestyle changes, such as quitting college or deciding not to marry, even though the disorder may never develop.

The shared nature of genetic information means that one relative's decision to undergo testing impacts the entire family. For example, upon learning that Uncle Joe has the gene for Huntington's disease, his family members may feel compelled to undergo testing before they decide to marry or have children to avoid passing along the dreaded gene.[423] In effect, Uncle Joe's decision to be tested can potentially create a "family curse," and the stigma surrounding his disease will cloud the entire family. An employer who sees Huntington's disease listed in an employee's family history may discriminate against her even though she has not been tested and may not carry the gene. Not surprisingly, by some estimates, fewer than 14% of persons at risk for Huntington's disease opt to undergo testing.[424] Clearly then, genetic testing has implications for Uncle Joe and his entire family. Therefore, shouldn't a person have the right *not* to know his or her genetic information?

Raising further privacy concerns is the realization that once a DNA sample is taken, the biological product is deemed "abandoned" under most state laws, meaning that the person has extinguished any property interest in the donated DNA. In *Moore v. Regents of University of California*, Moore

---

L. & POL'Y 365, 370–76 (2000) (discussing a study of 388 persons, in which 36% of those who tested positive for the BRCA-1 gene experienced sadness, 6% experienced anger, and 8% felt guilty).

419. Lee Bowman, *Many Don't Want Results from Genetic Tests*, CLEVELAND PLAIN DEALER, June 26, 1996, at 1A.

420. 2 ENCYCLOPEDIA OF BIOETHICS, *supra* note 27, at 953.

421. Weaver, *supra* note 233, at 253–57.

422. Kass, *supra* note 95, at 42 (telling the story of Prometheus who gave man the gift of "blind hopes" knowing "that ignorance of one's own future fate was indispensable to aspiration and achievement").

423. *See* 2 ENCYCLOPEDIA OF BIOETHICS, *supra* note 27, at 953 (discussing the ethics phenomenon of "relational responsibility").

424. Andrews, *supra* note 13, at 45.

underwent treatment for hairy cell leukemia.[425] Physicians decided to remove his spleen to use it for research but allegedly never informed Moore of their intent.[426] After his spleenectomy, Moore returned for medical visits on doctor's orders, and the doctor took more tissue samples.[427] All the while, Moore's doctors were benefiting financially and competitively from exclusive access to his tissues.[428] They used Moore's T-lymphocytes to establish a cell line from which they accrued substantial royalties and obtained a patent that also covers various methods for using the cell line to produce lymphokines.[429] Despite his physicians' behavior, the court noted that Moore stated a cause of action for breach of fiduciary duty and lack of informed consent but not for conversion.[430] The court stated that no person has a continuing proprietary interest in his cells once he has consented to their removal from his body and noted that to hold otherwise would hinder scientific research and product development.[431]

While extinguishing a property interest does not necessarily extinguish one's privacy interest in donated tissue, precedent suggests that just like trash[432] or excrement,[433] abandoned tissue may not be embraced by the Fourth Amendment. Arguably then, a person has no protectable privacy interest in the genetic information to which that abandoned tissue gives rise. This is especially alarming when one considers that DNA may be obtained "from the tip of a cigarette or the rim of a glass," a single strand of hair that falls unnoticed onto a keyboard, or even sloughed-off skin cells so minute as to be virtually invisible to the naked eye.[434]

Perhaps it is no wonder then why two employees of a Boston telecommunications company allege that they refused to give hair samples for drug testing and medical research but offered to provide urine samples. Still, the employees were terminated from their employment.[435] The employees feared that their hair would be kept in a plastic bag, labeled with

---

425. Moore v. Regents of Univ. of Cal., 793 P.2d 479, 481 (Cal. 1990), *cert. denied*, 499 U.S. 936 (1991).

426. *Id.*

427. *Id.*

428. *Id.*

429. *Id.* at 481–82.

430. *Id.* at 497.

431. *Id.* at 487–91; *see* Yesley, *supra* note 347, at 664 (rejecting the argument that John Moore had a property interest in cells removed from his body and later used without his permission).

432. *See* California v. Greenwood, 486 U.S. 35, 37 (1988).

433. Venner v. State, 354 A.2d 483 (1976) (ruling that a man abandoned his excrement for purposes of the Fourth Amendment).

434. Helena Kennedy, *Bing's Genes Concern Us All*, THE GUARDIAN (London), May 22, 2002, at 17. Without Steve Bing's consent, a private investigator allegedly removed Bing's dental floss from his dustbin and genetically analyzed it to prove that Bing was the biological father of a child of a well-known athlete married to a movie star. *Id.*

435. Shoop, *supra* note 19, at 12, 13.

their social security numbers, and be used later for genetic testing.[436] In 1997, they filed a wrongful termination suit on grounds that their employer had refused to answer their questions about genetic testing and had violated their rights to privacy as well as their Fourth Amendment right to freedom from unreasonable search and seizure.[437]

In another case, when two U.S. Marines refused to give blood for purposes of creating a national DNA databank for American soldiers because of their fear of potential genetic discrimination, they were charged with insubordination and launched into the national spotlight.[438] The government defended its DNA collection program on the grounds that it could be used to identify soldiers' remains.[439] The Ninth Circuit dismissed the claim as moot because the soldiers were honorably discharged while it was pending.[440]

The Fourth Amendment provides some comfort to anxious government employees. In *O'Connor v. Ortega*,[441] the United States Supreme Court held that a government employee had a reasonable expectation of privacy in his locked desk and filing cabinet, and in a later case, the Court held that a drug test was a "search" within the meaning of the Fourth Amendment.[442] However, *Ortega* provided insufficient protection for two reasons: (i) it applied only to government employees, and (ii) it only required that a government "search" be reasonable and lawful.[443]

Not surprisingly, employees largely supported GINA's enactment because they were uncomfortable with the fact that their abandoned DNA could be retained by their employers indefinitely and could be disclosed as outside the traditional, privileged doctor–patient relationship.[444] This potential lack of confidentiality could have been exacerbated as employers increasingly turned to independent contractors and labs to conduct genetic tests and to provide screening data.[445] If a workers' compensation suit arose, a person's entire genetic record, including tests for conditions unrelated to the alleged workplace injury or occupational disease, could be disclosed to the employer's insurers, management, and general counsel,

---

436. *Id.*

437. *Id.*

438. George Rodrigue, *Marine's Case Raises Medical Privacy Issues; He Refused To Give DNA Sample, Fearing Potential Discrimination*, DALL. MORNING NEWS, Apr. 12, 1996, at 1A.

439. Pagnatarro, *supra* note 199, at 144.

440. *Id.*

441. 480 U.S. 709, 719 (1987).

442. Kaufmann, *supra* note 239, at 431 (discussing Skinner v. Ry. Labor Executives' Ass'n, 489 U.S. 602, 633 (1989)).

443. *See Ortega*, 380 U.S. at 723.

444. Draper, *supra* note 367, at 294.

445. *Id.*

often without the employee's knowledge or consent.[446] It is no wonder then why genetic testing in employment has raised so many privacy concerns.[447]

GINA addresses these privacy concerns by prohibiting the disclosure of genetic information, except under limited circumstances,[448] and by requiring covered entities to maintain genetic information on a separate form in a separate medical file marked "confidential" and to treat the information as a confidential medical record.[449] More importantly, it prohibits covered entities from requesting, requiring, or purchasing genetic information or discriminating on the basis of that information.[450]

V. LEGISLATIVE SPLICING: IMPROVING TITLE II OF GINA

GINA is necessary to prohibit genetic discrimination in employment and has been drafted to adequately address many of the criticisms that have been lodged against it. That being said, GINA still warrants improvement. Thus, the remainder of this Article will highlight issues to consider in future amendments to the law and its accompanying regulations.

A. Expand the Definitions of "Genetic Information" and "Genetic Test"

The definitions of "genetic information"[451] and "genetic test"[452] should include the following: (i) mitochondrial DNA; (ii) genes; (iii) gene variants; (iv) genetic markers indicating a predisposition to any genetic disorder, disease, or condition; and (v) genes or genetic markers indicating carrier status of a gene indicative of a genetic disorder, disease, condition, or predisposition for any of the above. This will clarify that that carriers and predisposed persons also come within the Act's embrace. Furthermore, the regulations accompanying GINA should elucidate the role of "gene products" which indicate whether the individual actually has the genetic condition.[453]

---

446. *Id.*

447. *Id.* (analogizing genetic screening to existing drug screening policies and arguing that both screening measures identify problem employees while raising employees' privacy concerns).

448. *See supra* note 181 and accompanying text.

449. *See supra* note 178 and accompanying text.

450. *See supra* notes 106, 146, and accompanying text

451. *See supra* notes 108–12 and accompanying text.

452. *See supra* notes 122–27 and accompanying text.

453. Jonathan Riskind, *Include Gene Products in Bill on Test Prohibitions Says Expert*, COLUMBUS DISPATCH, Apr. 23, 1993, at 3C.

B. Rephrase Section 2(3)

Section 2(3) of GINA should be amended to change "sickle cell anemia, a disease which afflicts African-Americans"[454] to "sickle cell anemia, a disease that has traditionally been more prevalent among African-Americans than other racial and ethnic groups." The current statutory language erroneously suggests that sickle cell anemia is unique to African-Americans, thus perpetuating the genetic stigma that sickle cell anemia is the "black man's disease."[455]

C. Clarify Prohibited Conduct

All relevant sections of GINA should be amended or at least interpreted to prohibit employers from discriminating against applicants or employees because of genetic information, because of the covered individual's refusal to undergo genetic testing, *and* because of the individual's *refusal to disclose* genetic information, family history, or the results of any genetic tests to a covered entity.

D. Improve Confidentiality by Sealing Records in GINA Cases

If a person is required to submit genetic test results in order to state a claim, this might dissuade him or her from coming forward for fear of widespread disclosure of the genetic information. To remedy this issue, courts handling GINA cases should establish procedures to keep genetic discrimination cases under seal and, thus, truly confidential. As with vaccine injury cases litigated in the United States Court of Federal Claims, courts should also permit parties to propose redactions to sensitive information before any judicial opinion is published.

E. Clarify GINA's Relationship to Other Federal Legislation

GINA should clarify whether a suit brought under it may be concurrent with a Title VII, Section 1981, or other federal lawsuit arising from the same set of facts. For example, if an employer only tested African-Americans for sickle cell carrier status in violation of GINA, can those African-Americans bring a disparate impact race claim under Title VII in conjunction with individual disparate treatment GINA claims? While GINA clarifies that a person cannot bring suit under GINA and the ADA arising from the same set of facts, it does not adequately clarify GINA's relationship to other civil rights legislation.

---

454. Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-233, § 2(3), 122 Stat. 881, 882 (2008) (codified as amended at 42 U.S.C. §§ 2000ff to 2000ff-12 (Supp. III 2009)).

455. *See id.*

F. Clarify Whether GINA Applies Extraterritorially

In *EEOC v. Arabian Am. Oil Co.* ("*ARAMCO*"), the Supreme Court held that Title VII did not apply extraterritorially to regulate the employment practices of American employers that employ American citizens abroad.[456] The regulations accompanying GINA should incorporate *ARAMCO* to clarify GINA's extraterritorial application (*i.e.*, whether it applies to American companies operating in foreign countries, to American citizens working for American companies abroad, and to non-citizens working for American companies abroad).

G. Include a More Comprehensive Sunset Provision

At the conclusion of the current six-year sunset period, which requires establishment of a commission to make recommendations to Congress regarding whether a disparate impact cause of action should be available under GINA,[457] lawmakers should include a provision, mandating a review of the law every three years after its enactment. Considering the lightning pace of genetic and scientific breakthroughs as well as technological advances, it is extremely important that the entire statute be reviewed often for improvements, edits, and updates that keep pace with changing technology. Given the multifaceted implications of the law, the interdisciplinary reviewing commission should contain at least one of each of the following professionals: (i) a bioethicist; (ii) an attorney with expertise in genetic and health law; (iii) a geneticist; (iv) a physician; (v) a representative from the insurance industry; (vi) a member of Congress who sponsored GINA; (vii) a representative from the business community; (viii) a genetic researcher; (ix) a representative from the genetic testing industry; (x) a fertility specialist; (xi) a genetic counselor; (xii) a human resources director or specialist; (xiii) an employment attorney; and (xiv) a representative from the law enforcement community.[458]

H. Adopt an Informed Consent Model

Congress should consider changing Section 202(b)(2)(B)'s standard of "prior, knowing, voluntary, and written authorization" to "voluntary informed consent in writing." Informed consent is a standard often used in healthcare legislation and tort litigation. As such, the law of informed consent is well-developed and could assist courts in more easily

456. *EEOC v. Arabian Am. Oil Co.*, 499 U.S. 244 (1991).

457. 42 U.S.C. § 2000ff-7 (Supp. III 2009) (creating a commission to make recommendations to Congress regarding whether to provide a disparate impact cause of action under GINA).

458. *See* 42 U.S.C. § 2000ff-7(c) (Supp. III 2009) (listing who will appoint members of the commission but providing no insight on whom should be appointed).

determining whether GINA has been violated. "Voluntary informed consent in writing" should be required prior to any genetic testing and prior to each and all uses, releases, and disclosures of genetic information as defined by the Act, including both test results, for any reason whatsoever, and participation in federal research studies.

I. Clarify GINA's Impact on Workers' Compensation

The regulations accompanying GINA should specifically address whether GINA permits employers, upon the demand of a workers' compensation carrier, to request information about an employee's genetic predisposition when the employee suffers an occupational injury and seeks workers' compensation coverage. The reason for this is that many workers' compensation statutes have heightened causation prongs for pre-existing conditions or disorders. If an individual possesses a predisposition for cancer and that cancer is triggered by workplace exposure to carcinogens, critics could argue that the employee bears the heightened burden of proving that the workplace carcinogens are peculiar to or increased by the employment. Because the employee is making his health an issue, tort principles dictate allowing the employer to investigate whether or not the employee had such a genetic predisposition. However, it is somewhat unclear whether such an investigation or request would be permissible under GINA. Therefore, regulators or lawmakers should clarify whether in such cases, an employer or workers' compensation insurer may request, require, or use the claimant's genetic information to determine whether the individual falls under the state's heightened causation prong.

J. Add Lab Restriction

GINA should be amended to include a restriction on the use of commercial laboratories in conducting any permissible genetic tests. Such a provision might state:

> A laboratory receiving a request to conduct a genetic test or analysis of donated tissue may conduct the requested test or analysis only after receipt of a statement signed by the individual or entity lawfully ordering the test or analysis, which states and clearly confirms that the tissue donor gave his prior, knowing, and voluntary consent in writing to the performance of this genetic test or analysis on this donated tissue and that the request was made in full compliance with applicable federal, state, and local law and regulations.

K. Consider Allowing Disparate Impact Claims

GINA currently disallows disparate impact claims. Congress should consider amending GINA to permit disparate impact claims so that the

statute will be more consistent with other civil rights legislation, such as Title VII, which expressly permits disparate impact claims. Also, genetic discrimination can have a disparate impact because genes, though facially neutral, are "[often] associated with particular racial or ethnic groups and gender. Because some genetic traits are most prevalent in particular groups, members of a particular group may be stigmatized or discriminated against as a result of that genetic information."[459] The advent of sickle cell anemia programs to screen carriers in the 1970s illustrates this type of discrimination.

L. Consider Adding a BFOQ Provision

Congress should consider adding a bona fide occupational qualification ("BFOQ") prong to GINA, which would be narrowly construed and would ensure that a person whose genetic information genuinely disqualifies him or her from a position can be lawfully excluded. The provision might read:

> No applicant or present employee who is otherwise qualified for a position shall be denied equal opportunities in employment solely because of such person's genetic information, unless it can be clearly shown that the person's genetic information indicates a condition that would at present prevent such person from performing this particular job effectively and that would pose danger to the safety of such person or of others.

M. Clarify the Applicability of the Direct-Threat-to-Self Exception

Ecogenetic and epigenetic research reveals the interplay between genetic expression and environmental triggers.[460] For example, because PKU only manifests if the individual ingests phenylalanine, manifestation of PKU symptoms may be prevented or at least minimized when detected at birth.[461] Accordingly, GINA should clarify whether in cases where workplace triggers may provoke gene expression and thus, pose a threat to the individual's health or that of third parties, employers may take the genetic information into account when making employment decisions or when limiting, segregating, or classifying employees.

N. Address Genetic Propensity for Violence

Some studies in behavioral genetics indicate that "a genetic mutation in the structural gene for monoamine oxidase A, which causes an acute

---

459. Genetic Information Nondiscrimination Act of 2008 Pub. L. No. 110-233, § 2(3), 122 Stat. 881, 882 (2008) (codified as amended at 42 U.S.C. §§ 2000ff–2000ff-12 (Supp. III 2009)).

460. Lorber & Perdue, *supra* note 5, at 15.

461. *Id.*

buildup of neurotransmitters associated with 'fight or flight' stress responses" may be associated with males exhibiting abnormal behavior, including, but not limited to, impulsive aggression.[462] "At least one criminal defendant has already attempted to mitigate his sentence by undergoing a genetic test to determine whether he possesses the 'mean gene.'"[463] Prior to GINA, research uncovering genetic causes for aggression might have played an interesting role in cases alleging negligent hiring or supervision due to violence perpetrated by employees.[464] Now, however, even if such genes are discovered, GINA will prohibit employers from testing employees or from making employment decisions on the basis of the employees' possession of genes related to aggression, even where the employees work with children or in high stress situations that could trigger the gene's expression. Even if an employee knows that he possesses the mean gene or that such aggression runs in his family, under GINA, his employer cannot lawfully inquire about that information; without it, the employer cannot predict whether the applicant is genetically predisposed to violence and poses a risk to his safety or that of others.[465]

Ensuring GINA's success is not a responsibility that falls only to Congress and the regulators. Rather, to maximize GINA's effectiveness, educational and ad campaigns aimed at raising awareness of the new law are necessary.[466] Law professors should be encouraged to teach genetic discrimination in their employment discrimination courses, and law firms should offer continuing legal education courses to update lawyers on GINA and its accompanying regulations. Employers should incorporate the law into handbooks and training manuals. Human resources specialists should conduct training seminars for management and employees to explain the new law and to encourage compliance. The EEOC website should continue to provide information regarding basic genetic information underlying the law along with a thorough and clear explanation of the statute and its accompanying regulations.

---

462. *Id.* at 16.

463. *Id.*

464. *Id.*

465. *Id.*

466. *See, e.g.*, Mark A. Rothstein, Betsy D. Gelb & Steven G. Craig, *Protecting Genetic Privacy by Permitting Employer Access Only to Job-Related Employee Medical Information: Analysis of a Unique Minnesota Law*, 24 AM. J.L. & MED. 399, 410 (1998) (admitting that many employment lawyers, employers, and occupational physicians do not know about the current Minnesota law aimed at preventing genetic discrimination).

VI. CONCLUSION

Genetics will continue to play an increasingly pervasive role in our lives:

> [Genetic issues] promise[] . . . to alter the very nature of humanity. . . . [and] raise[] questions which we barely have a vocabulary to discuss, much less social and political institutions to decide. With power comes the responsibility of choosing wisely. . . . [which] depends on accurate information. . . . [W]e have come full circle, to a time when everyone must be a biologist, and the world is a classroom![467]

Our expanding understanding of and access to genetic information necessitated the evolution of the laws and regulations governing its use. As the most recent development in that natural and necessary legal evolution, GINA is a prophylactic measure that will better ensure that Americans may undergo beneficial genetic testing and treatment and participate in genetic research without endangering their privacy or ability to obtain insurance and employment.[468]

467. GONICK & WHEELIS, *supra* note 23, at 209.

468. Lorber & Perdue, *supra* note 5, at 17.

NO WALK IN THE DOG PARK: DRAFTING ANIMAL CRUELTY STATUTES TO RESOLVE DOUBLE JEOPARDY CONCERNS AND ELIMINATE UNFETTERED PROSECUTORIAL DISCRETION

LAURIE SERAFINO*

I. INTRODUCTION 1120
II. THE EVOLVING NATURE OF ANIMAL CRUELTY 1123
A. Historical Foundations 1123
B. Modern Statutes 1124
C. Areas of Particular Concern 1127
1. Cockfights and Dogfights 1127
2. Videos Showing Animals Being Abused 1128
3. Endangered Species 1129
4. Animal Hoarders 1130
5. Violent Offenders 1131
6. Acceptable Animal Cruelty 1133
III. ALTERNATIVES AND WHY RESOLUTION IS IMPORTANT 1136
A. Statutory Alternatives: Summary of the Variations 1136
1. Animal Cruelty Statutes Authorizing One Count per Animal 1137
2. The Statute is Silent 1138
3. A Separate Offense if the Violation Continues to Another Day 1139
4. Some Animals Are More Important than Others 1139
5. One Count per Animal if a Certain Type of Abuse or Abuser 1140
6. One Count per Animal in Oregon? 1141
B. Why Resolution of This Disparity and Ambiguity is Important 1142
IV. DOUBLE JEOPARDY AND THE UNIT OF PROSECUTION 1145
A. Ascertaining Units of Prosecution 1146

---

*Assistant Professor of Law, Pepperdine University School of Law. This work would not have been possible without the invaluable help of Pepperdine Law School Librarian Jennifer Allison. The author also wishes to thank Professor Chris Chambers Goodman for her mentoring and conscientious review, the Dean's Summer Research Grant Fund for providing financial support to this project, and Kristen Marquis, Amir Tikriti, and Catherine Moore for their research assistance. Finally, I thank my husband Stephen, and my children Abigail, Claire, and Christopher. All website references are current as of (Dec. 21, 2011).

B. The Struggle in Defining the Unit of Prosecution in Animal Abuse Cases 1149
C. Michigan's Model 1153
V. RECOMMENDATIONS 1155
A. The Human Victim, Property, Living Property, Personhood, or Equal Protection Model? 1155
B. Hypotheticals 1161
C. What About the Michigan Model? 1162
D. A Proposed Statute 1164
V. CONCLUSIONS 1165

I. INTRODUCTION

Over the past few decades, animal protection laws enacted across the country have become more voluminous, broad, and complex.[1] Our laws reflect a movement from the historical view that the owners of animals, due to economic and personal loss, were the victims of animal abuse to a belief that the animals themselves are the harmed parties.[2] The desire is emerging to cause as little harm as possible to animals and to "render aid to relatively vulnerable and helpless animals when faced with people willing or even anxious to mistreat them."[3]

At the same time, as highlighted by the United States Supreme Court in *United States v. Stevens*, each state defines "animal" and "animal cruelty"

---

1. *See* Stephan K. Otto, *State Animal Protection Laws—The Next Generation*, 11 ANIMAL L. 131, 134 (2005) ("Since 1990, thirty-five states and the District of Columbia have enacted, for the first time, felony-level laws for certain types of animal abuse.").

2. *See generally* Luis E. Chiesa, *Why Is It a Crime to Stomp on a Goldfish?—Harm, Victimhood and the Structure of Anti-Cruelty Offenses*, 78 MISS. L.J. 1 (2008), *available at* http://www.animallaw.info/articles/arus78misslj1.htm (last visited Dec. 21, 2011) (discussing the purpose and history of animal cruelty laws). Some of the anti-cruelty laws initially enacted in this country made it a criminal offense to engage in abusive acts against animals only if they were owned by another person. *Id.* at 8 (citing 1846 VT. ACTS & RESOLVES 34). However, most states have now amended their statutes to protect the animals themselves. *See, e.g.*, LA. REV. STAT. ANN. § 14:102.1(4) (2011) ("Any person who intentionally or with criminal negligence mistreats any living animal whether belonging to himself or another by any act or omission which causes or permits unnecessary or unjustifiable physical pain, suffering, or death to the animal shall also be guilty of aggravated cruelty to animals."); *see also* ALA. CODE § 2-15-110 (2010) ("In order to prevent injury to animals in livestock markets and in transit and to prevent unnecessary abuse and cruelty to animals with resultant loss of profit from the slaughter and sale of such animals, it shall be unlawful in this state to handle or transport such animals in any manner not consistent with humane methods of treatment to such extent as is reasonably possible . . . .").

3. State v. Bauer, 379 N.W.2d 895, 899 (Wis. Ct. App. 1985).

in its own way.[4] A "bewildering maze of regulations from at least 56 separate jurisdictions" presents law enforcement officials with the difficult task of determining which acts against what animals are unlawful.[5] At the most basic level, state laws cannot agree on even a common definition of "animal." States protect animals by defining them in a myriad of ways. In Kansas, abuse against animals is defined as a "nonperson" offense.[6] In Ohio, dogs are considered "victims," and animals analogized to "humans."[7] Tennessee refers to "victimized animals,"[8] and Michigan and Utah refer to animals as nonhuman, vertebrate creatures.[9] And finally, the District of Columbia defines animals as "all living and sentient creatures (human beings excepted)."[10] Some state statutes and case law reflect an underlying belief that animals are entitled to the protection of the law for more than their property status or worth in a non-criminal law context.[11]

Similarly, the term "animal cruelty" may describe a broad range of mistreatment, from an owner's temporary lapse in providing proper care to the malicious torture or killing of an animal.[12] For example, in Arizona, animal cruelty is not a crime of moral turpitude that would entitle a defendant to a jury trial;[13] yet, in Mississippi, any person who maliciously, either out of a spirit of revenge or wanton cruelty, kills, maims, wounds, or injures any livestock is guilty of a felony;[14] still, in Tennessee, the intentional killing of an animal is a crime only if it is accomplished without the owner's consent and constitutes theft of property, graded according to the animal's value.[15]

---

4. United States v. Stevens, 130 S. Ct. 1577, 1589 (2010).

5. *Id.* at 1589.

6. KAN. STAT. ANN. § 21-4310(d) (Supp. 2010).

7. *See* State v. Angus, No. 05AP-1054, 2006 WL 2474512, at *7 (Ohio Ct. App. Aug. 29, 2006) (explaining that a defendant accused of harming two dogs had "committed crimes that involved *two separate victims*").

8. State v. Webb, 130 S.W.3d 799, 836–37 (Tenn. Crim. App. 2003).

9. MICH. COMP. LAWS ANN. § 750.50(1)(b) (2011); UTAH CODE ANN. § 76-9-301(b)(i) (2011).

10. D.C. CODE § 22-1013 (2011).

11. DAVID FAVRE, ANIMAL LAW WELFARE, INTERESTS, AND RIGHTS 431 (2008) [hereinafter ANIMAL LAW WELFARE] ("[A]nimals are presently categorized and treated as property in our legal system.").

12. *See* M. Varn Chandola, *Dissecting American Animal Protection Law: Healing the Wounds with Animal Rights and Eastern Enlightenment*, 8 WIS. ENVTL. L.J. 3, 5 (2002) (describing animal cruelty broadly using Colorado's statutory definition).

13. Campbell v. Superior Court, 924 P.2d 1045, 1046 (Ariz. Ct. App. 1996). Generally, a crime of moral turpitude is an act of baseness, vileness, or depravity. State v. Malusky, 230 N.W. 735, 737 (N.D. 1930).

14. MISS. CODE ANN. § 97-41-15 (2010). A malicious, criminal act is one which is "naturally evil." *See* State v. Horton, 51 S.E. 945, 946 (N.C. 1905).

15. TENN. CODE ANN. § 39-14-205(a)(1) (2010).

This article demonstrates that a growing trend treats animals more like humans and less like property in criminal cases and statutes. Yet, the laws of many states still adhere to the traditional view of animals as property, causing unique charging and sentencing issues that must be clarified in order to bring predictability and consistency to the law. This article considers whether animals, which, as Justice Alito stated, are "living creatures that experience excruciating pain,"[16] should be treated like human criminal victims or property or whether a new paradigm should be created that treats animals as a hybrid category of "living property," "legal personhood," or some other type of entity entitled to equal protection.[17]

This article also addresses whether the injury of each and every animal can or should constitute a separate offense, as well as the extent to which double jeopardy principles apply to bar sentences on multiple counts.[18] Few states have legislated the issue of charging defendants when multiple animals are injured at the same time. This article will demonstrate that, by failing to legislate in this area, forty-one states are in effect allowing prosecutors unfettered discretion to charge defendants in these cases in whatever way they choose.[19] Without legislation to guide them, prosecutors decide whether each harmed animal should constitute a separate unit of prosecution or whether all injured or dead animals, regardless of their number, should be joined together in one count.[20] This article shows that it is imprudent to leave such important charging decisions to prosecutors and that clear, uniform, and organized charging and sentencing statutes eliminate disparity, provide guidance to prosecutors, and protect the constitutional rights of criminal defendants.

When animal abuse cases involving multiple animals are appealed, state courts are left to address the constitutionality of the resulting convictions. State courts are resolving these Fifth Amendment double jeopardy issues in a variety of ways.[21] Some courts are applying rules usually invoked in crimes against property and are concluding that each incident of abuse against multiple animals constitutes one criminal transaction.[22] Courts in this group hold that defendants cannot receive

16. United States v. Stevens, 130 S. Ct. 1577, 1600 (2010).

17. *See infra* Part V.

18. Breed v. Jones, 421 U.S. 519, 537 (1975) (explaining that it is a fundamental principle of our constitutional system that a defendant may not be placed twice in jeopardy for the same offense and that the double jeopardy clause protects against multiple punishments for the same offense); Will v. United States, 389 U.S. 90, 98 (1967). "The Fifth Amendment's prohibition against placing a defendant 'twice in jeopardy' represents a constitutional policy of finality for a defendant's benefit in federal criminal proceedings." United States v. Jorn, 400 U.S. 470, 479 (1971).

19. *See infra* Part III.

20. *See infra* Part III.

21. *See infra* Part IV.

22. *See infra* Part III.

multiple consecutive sentences for each abused animal, deeming such outcomes multiple punishments for the same offense at one trial.[23] Other courts analogize animals to human victims.[24] These courts conclude that where a defendant's conduct involves multiple "victims," the defendant should be convicted and sentenced for each injured animal.[25]

To provide context, this article provides an overview in Part II of how animal cruelty statutes have evolved from laws designed to protect property and to enforce moral character into genuine efforts to protect animals from harm. Part II considers charging and sentencing issues that emerge in prosecuting cockfight and dogfight participants, animal abuse video producers, those who threaten endangered species, animal hoarders, and violent offenders. It also examines charging and sentencing issues emerging in what are normally legally acceptable animal abuse situations like hunting, farming, and scientific research that, for various reasons, cross the line and become criminal actions.

Part III explores the challenges faced by courts and legislatures as they grapple with situations of abuse involving multiple animals and gives examples of the ways states are handling this issue. Part III also explains why the resolution of disparities and ambiguities in animal abuse statutes is important. Part IV addresses the concept of double jeopardy as applied to multiple human victim and property cases and considers the proper unit of prosecution in animal abuse cases. It also looks at Michigan's unique statutory scheme. Finally, Part V analyzes the pros and cons of various charging and sentencing options and proposes a statutory scheme that ensures fair, just treatment.

II. THE EVOLVING NATURE OF ANIMAL CRUELTY

A. Historical Foundations

"Expanding urbanization in the Victorian era brought about a shift in the way humans saw other animals. 'Victorians no longer viewed animals as commodities or tools, but as companions and even members of the family.'"[26] During this period, states extended protections to animals as a

---

23. *See infra* Part III.

24. *See infra* Part III.A.1.

25. *See infra* Part III.A.1.

26. Corwin R. Kruse, *Baby Steps: Minnesota Raises Certain Forms of Animal Cruelty To Felony Status*, 28 WM. MITCHELL L. REV. 1649, 1654 (2002); Charles M. Friend, *Animal Cruelty Laws: The Case for Reform*, 8 U. RICH. L. REV. 201, 201–02 (1974) (discussing the concept of animals as property). *See generally* GARY L. FRANCIONE, ANIMALS, PROPERTY, AND THE LAW (1995) (providing a useful overview of animal abuse law); David Favre & Vivian Tsang, *The Development of Anti-Cruelty Laws During the 1800's*, 1993 DET. C.L. REV. 1 (1993) [hereinafter Favre, *Development*] (discussing the history of modern anti-cruelty legislation).

means of enforcing moral character and obtaining reimbursement for damages when animals were harmed.[27] In 1867, a New York statute extended its reach to "any living creature," thereby eliminating the principle that protection was intended only for animals of commercial value.[28] The New York statute also removed the *mens rea* requirement from all of the prohibited acts,[29] marking a shift in focus from the subjective mindset of the accused to the objective harm to the animal.

Fast-forward to 2010, when Justice Alito, analyzing the constitutionality of a statute regulating videos depicting animal cruelty, stated that the most relevant prior decisions of the United States Supreme Court concerned child pornography.[30] According to Justice Alito, child pornography "involves the commission of a crime that inflicts severe personal injury to the 'children who are made to engage in sexual conduct for commercial purposes.'"[31] In comparison, a video depicting animal cruelty "records the actual commission of a criminal act that inflicts severe physical injury and excruciating pain and ultimately results in [an animal's] death."[32] Although Justice Alito acknowledged that an animal protection law "differs from a child pornography law in an important respect: preventing the abuse of children is certainly much more important than preventing the torture of the animals,"[33] his statement is highly significant. His comment reflects the view of most people that a civilized society has an interest in keeping animals free from unnecessary suffering and the unjustified infliction of pain.[34]

B. Modern Statutes

So long as anti-cruelty statutes are conceived of as laws protecting animals from enduring direct suffering, they are not property-related "victimless" crimes. The legally protected victim is the animal harmed by the perpetrator's conduct. For example, state laws protect animals not owned by anyone, reflected by the fact that torturing stray dogs is a felony in many states.[35] Dogfights and cockfights involve creatures that are not

---

27. *See generally* Favre, *Development*, *supra* note 26 (discussing the history of animal cruelty laws and their purposes).

28. *Id.* at 15–17.

29. *Id.*

30. United States v. Stevens, 130 S. Ct. 1577, 1599 (2010) (Alito, J., dissenting) (citing New York v. Ferber, 458 U.S. 747 (1982)).

31. *Id.* at 1599 (citing *Feber*, 458 U.S. at 753).

32. *Id.*

33. *Id.* at 1600.

34. *See, e.g.*, WASH. REV. CODE § 16.52.207(1) (2010) ("A person is guilty of animal cruelty in the second degree if, under circumstances not amounting to first degree animal cruelty, the person knowingly, recklessly, or with criminal negligence inflicts unnecessary suffering or pain upon an animal.").

35. Chiesa, *supra* note 2, at 25.

usually owned by the producers of the fights;[36] yet, animal cruelty laws protect these creatures.[37] Furthermore, every jurisdiction now makes it criminal for a pet owner to mistreat his or her own animals.[38] These statutes are directly at odds with a property-based concept of animal ownership in which an owner has an absolute right to do as he wishes with his property, including destroying or damaging it.[39]

Modern animal abuse statutes are also more than a means of protecting people from emotional harm. Animals are deemed worthy of legal protection whether or not they are generally liked—creatures like snakes and hamsters that do not normally have a close, daily relationship with humans are still protected.[40] Moreover, harming stray dogs or cats is a crime even if no one has developed a strong emotional bond with the animals.[41] Similarly, dogfights and cockfights are criminally sanctioned, despite the fact that spectators and promoters do not suffer when the animals are in pain, but instead enjoy watching the suffering.[42]

Animal cruelty laws also serve a greater purpose than predicting future harm to humans. This is evident from the criminalization of the negligent mistreatment of pets and the prohibition of dogfights and cockfights.[43] For instance, no evidence exists that the negligent treatment of animals or the promotion of dogfights increases the risk that the perpetrator will later abuse humans. Nevertheless, although such conduct has not been shown to have a correlation between spousal, child, and elder abuse,[44] it is sanctioned for its harmful consequences in and of themselves.

---

36. *Id.* at 10.

37. *See, e.g.*, ARIZ. REV. STAT. ANN. § 13-2910.03(A)(2) (2011) ("A person commits cockfighting by knowingly . . . for amusement or gain, causing any cock to fight with another cock or causing any cocks to injure each other.").

38. Chiesa, *supra* note 2, at 27.

39. *Id.*; *see also* Am. Sheet & Tin Plate Co. v. Pittsburgh & L.E.R. Co., 143 F. 789, 793 (3d Cir. 1906) ("The exclusive control of private property is subordinate to the exigencies of public safety and private necessity.").

40. Chiesa, *supra* note 2, at 30.

41. *Id.*

42. *See id.* at 30; *see, e.g.*, CAL. PENAL CODE § 597(c) (West 2011). Several animal cruelty statutes make it a crime to harm any mammal, bird, reptile, amphibian, or fish. *See, e.g.*, TEX. PENAL CODE ANN. § 42.092(a)(2) (West 2010) ("'Animal' means a domesticated living creature, including any stray or feral cat or dog, and a wild living creature previously captured.").

43. *See* CAL. PENAL CODE § 597.5 (West 2011) (explaining that a person shall not engage in animal fighting); MICH. COMP. LAWS ANN. § 750.49 (West 2011) (explaining that a person shall not be a party to animal fighting); MICH. COMP. LAWS ANN. § 750.50(2)(a) (West 2011) ("An owner, possessor, or person having the charge or custody of an animal shall not . . . fail to provide an animal with adequate care.").

44. *See* STEPHEN R. KELLERT & ALAN R. FELTHOUS, *Childhood Cruelty Toward Animals Among Criminals and Non-Criminals*, *in* CRUELTY TO ANIMALS AND INTERPERSONAL VIOLENCE 194, 208 (Randall Lockwood & Frank R. Ascione eds., 1998)

Animal cruelty statutes are no longer simply a means of enforcing a moral principle.[45] We have moved beyond the place where, as the Model Penal Code states, "the object of such statutes [is] to prevent outrage to the sensibilities of the community."[46] Animal laws now regulate conduct that was formerly accepted or even valued in our society, such as dyeing rabbits and chicks during Easter[47] or organizing cockfights and dogfights.[48]

In response to this societal shift in how animals are viewed,[49] the federal government and state legislatures continue to seek ways to prevent animals from needless suffering.[50] Statutes, for example, regulate the use of gas chambers for companion animals and require humane killings.[51] In Michigan, before an abused animal can be humanely euthanized, a hearing must be held to determine whether it lacks any useful purpose or whether it constitutes a public safety threat.[52] Additionally, Wisconsin forbids cruelty to rodents except when the poison is used on one's own premises.[53]

(examining the relationship between childhood cruelty toward animals and behavior among criminals and non-criminals in adulthood).

45. Chiesa, *supra* note 2, at 35–37. *But see* Chandola, *supra* note 12, at 30 ("[T]he law, instead of recognizing the intrinsic worth of animals, only concerns itself with protecting public morals. Even when focusing on human conduct, institutionalized uses of animals for food, clothing, science, and entertainment are not considered immoral.").

46. MODEL PENAL CODE § 250.11 (1962) ("A person commits a misdemeanor if he purposely or recklessly: (1) subjects any animal to cruel mistreatment; or (2) subjects any animal in his custody to cruel neglect; or (3) kills or injures any animal belonging to another without legal privilege or consent of the owner. Subsections (1) and (2) shall not be deemed applicable to accepted veterinary practices and activities carried on for scientific research.").

47. *See, e.g.*, S.C. CODE ANN. § 47-1-125(1) (2010) ("It is unlawful for any person to dye or color artificially any animal or fowl, including but not limited to rabbits, baby chickens, and ducklings, or to bring any dyed or colored animal or fowl into this State."); *see also* R.I. GEN. LAWS § 4-1-8 (2010).

48. *See, e.g.*, N.M. STAT. ANN. § 30-18-9(A) (West 2011) ("It is unlawful for any person to cause, sponsor, arrange, hold or participate in a fight between dogs or cocks for the purpose of monetary gain or entertainment.").

49. In another example of the changing attitudes towards animals, companies now brag that their products are not tested on animals. *See, e.g.*, *Our Commitment to Natural Products*, BURT'S BEES, http://www.burtsbees.com/c/commitment/; *Frequently Asked Questions*, HERBAL ESSENCES, http://www.herbalessences.com/en-us/frequently-asked-questions; *Not Tested on Animals*, CARMEX, http://www.mycarmex.com/never-tested-on-animals/default.aspx.

50. *See* Tom Breen, *USDA Found Problems at NC Animal Research Lab*, THE BOSTON GLOBE, Oct. 8, 2010, *available at* http://www.boston.com/news/nation/articles/2010/10/08/usda_found_problems_at_nc_animal_research_lab/.

51. *See, e.g.*, VA. CODE ANN. § 3.2-6505 (2011).

52. MICH. COMP. LAWS ANN. § 750.49(17) (West 2011); *see also* 510 ILL. COMP. STAT. ANN. 70/3.09 (West 2011).

53. WIS. STAT. ANN. § 951.06 (West 2010).

Furthermore, many consumers are demanding free-range, organic meat and eggs.[54]

C. Areas of Particular Concern

In light of these important changes in the perception of animals in our society, a review of some types of animal abuse that present unique charging and sentencing issues is appropriate.

1. Cockfights and Dogfights

Although Washington, Jefferson, and Hamilton were devotees of cockfighting, and Lincoln, who umpired cockfights, stated, "[a]s long as the Almighty permitted intelligent men, created in His image and likeness, to fight in public and kill each other while the world looks on approvingly, it's not for me to deprive the chickens of the same privilege,"[55] cockfights and dogfights are now universally condemned.[56] Now, like the court in *People v. Baniqued*, most of us are more likely to sympathize with the observation of another early American, Mark Twain, that "[man] is the only creature that inflicts pain for sport, knowing it to *be* pain."[57]

Nevertheless, despite law clearly condemning such conduct, courts have rarely considered charging and sentencing issues in cockfighting and dog fighting cases. In the absence of established law, many questions remain. For example, if a perpetrator is arrested for promoting dogfights, can he or she be charged and sentenced for each animal involved in the fighting, or is only one count encompassing the entire fight authorized? Can a perpetrator who is charged with the crime of promoting, engaging, or watching the fighting also be charged with the crime of raising the creatures for fighting? If the latter is true, can the perpetrator be held responsible for the injuries to each animal that is raised for fighting?

---

54. *See* League of Women Voters of California Education Fund, *Proposition 2—Standards for Confining Farm Animals—State of California*, (2008), http://www.smartvoter.org/2008/11/04/ca/state/prop/2/ ("Proposition 2 is a moderate measure that stops cruel and inhumane treatment of animals—ending the practice of cramming farm animals into cages so small the animals can't even turn around or stretch their limbs. Voting YES on Proposition 2 prevents animal cruelty, promotes food safety, supports family farmers, and protects the environment. The agribusiness interests opposing Proposition 2—masquerading as the deceptively named Californians for Safe Food—have a record of duping the public, harming animals, and polluting the environment.").

55. State v. Claiborne, 505 P.2d 732, 733 (Kan. 1973).

56. *See* 510 ILL. COMP. STAT. ANN. 70/4.01 (West 2011) (creating a duty to report for veterinarians presented with animals with wounds consistent with fighting); MISS. CODE ANN. § 97-41-19 (2010) (deeming it is a felony to be a *spectator* at a dogfight).

57. People v. Baniqued, 101 Cal. Rptr. 2d 835, 837 (Cal. Ct. App. 2000) (quoting 2 MARK TWAIN, MARK TWAIN'S AUTOBIOGRAPHY 7 (1924)).

Multiple jurisdictional issues must also be considered. For example, National Football League quarterback Michael Vick not only pleaded guilty to a federal dog fighting conspiracy charge, but also faced the following state felony dogfight charges in Virginia: one count each of beating, killing, or causing dogs to fight, and engaging in or promoting dogfights.[58] In 2008, Vick pleaded guilty in state court to dog fighting, and the remaining cruelty to animals charge was dismissed.[59]

2. Videos Showing Animals Being Abused

Crush videos depict "women slowly crushing animals to death 'with their bare feet or while wearing high heeled shoes,' sometimes while 'talking to the animals in a kind of dominatrix patter' over '[t]he cries and squeals of the animals, obviously in great pain.'"[60] Dogfight and cockfight videos record the actual commission of a crime involving deadly violence or resulting in suffering lasting "for years rather than minutes."[61]

United States v. Stevens addressed the constitutionality of a federal statute banning crush videos that was enacted in 1999.[62] The Supreme Court struck down this statute, holding that it was too broadly written and violated free-speech protections.[63] In 2010, a new measure was passed by Congress and signed into law by President Obama, which makes it a crime to sell or distribute videos that violate bans on animal cruelty by depicting "actual conduct in which 1 or more living [animals] is intentionally

---

58. *Vick, Michael-Associated Materials* (2007, 2008), MICHIGAN STATE UNIVERSITY COLLEGE OF LAW, ANIMAL LEGAL & HISTORICAL CENTER, http://www.animallaw.info/pleadings/pbusfdvick.htm (last visited Dec. 21, 2011).

59. A report prepared by the USDA's inspector general-investigations division stated that in 2007, quarterback Michael Vick and two co-defendants hung approximately three dogs that did not perform well in a "rolling session," which indicates the readiness of a dog to fight. Kelly Naqi, *In Virginia Facing State Dogfighting Charges, Vick's Involvement Revealed*, ESPN.com (Nov. 22, 2008, 11:34 AM), http://sports.espn.go.com/nfl/news/story?id=3718304. According to the report, the three men hung the dogs "by placing a nylon cord over a 2 X 4 that was nailed to two trees located next to the big shed. They also drowned approximately three dogs by putting the dogs' heads in a five gallon bucket of water." *Id.* Vick was sentenced to twenty-three months in prison for his role in the federal conspiracy and agreed to serve three years in prison and to pay a fine of $2,500. The three-year term and fine was suspended provided that Vick "remain[ed] of uniform good behavior for a term of (4) four years." Stipulation of Fact, Commonwealth v. Vick, No. CR-07-0000056 (Surry County (Va.) Cir. Ct., Nov. 2008), *available at* http://www.animallaw.info/pleadings/pb_pdf/pbusvavick_state_plea.pdf.

60. United States v. Stevens, 130 S. Ct. 1577, 1583 (2010) (alteration in original) (invalidated by 18 U.S.C. § 48 (2006)).

61. *Id.* at 1602.

62. *See id.*; *see also* 18 U.S.C. § 48 (2006).

63. *Stevens*, 130 S. Ct. at 1592.

crushed, burned, drowned, suffocated, impaled, or otherwise subjected to serious bodily injury."[64]

These videos, which show "extreme acts of animal cruelty that appeal to a specific sexual fetish,"[65] would seem to call for state and federal prosecutions.[66] Perpetrators could be prosecuted in state court for animal abuse and in state or federal court for the videos.[67] However, there are no cases considering the issue of whether prosecutions by both state and federal governments are barred by the constitutional protection against double jeopardy.[68]

3. Endangered Species

In 1973, the United States adopted the Endangered Species Act,[69] permitting criminal actions against private parties who cause harm to animals protected by the Act.[70] Criminal enforcement of this law often occurs in tandem with other federal laws.[71] Illegal conduct towards endangered species causes a ripple effect involving not only the endangered animal but other species as well.[72] Honking a horn at a cow is not a crime,

---

64. Animal Crush Video Prohibition Act of 2010, Pub. L. No. 111-294, § 48, 2010 U.S.C.C.A.N. (124 Stat.) 3177, 1378 (2010).

65. *Id.* at § 2.

66. For example, Illinois has its own statute banning depictions of animal cruelty. *See* 510 ILL. COMP. STAT. ANN. 70/3.03-1 (West 2010).

67. The problem is further complicated by the number of federal laws possibly invoked in animal cruelty cases. *See, e.g.*, Humane Methods of Slaughter Act of 1978, 7 U.S.C. § 1901 (2006); Animal Welfare Act of 1970, 7 U.S.C. § 2131 (2006) (specifically addressing interstate activities involving dogfights and cockfights); Bald and Golden Eagle Protection Act of 1940, 16 U.S.C. §§ 668–668(d) (2006); Wild Free-Roaming Horses and Burros Act of 1971, 16 U.S.C. §§ 1331–1340 (2006); Marine Mammal Protection Act of 1972, 16 USC §§ 1361–1421(h) (2006); Endangered Species Act of 1973, 16 U.S.C. §§ 1531–44 (2006); Wild Bird Conservation Act of 1992, 16 U.S.C. §§ 4901–16 (2006).

68. Federal and state governments have concurrent jurisdiction over much criminal activity. *See* 1 WAYNE R. LAFAVE & AUSTIN W. SCOTT, JR., SUBSTANTIVE CRIMINAL LAW 175–76 (1986); *see also* Bartkus v. Illinois, 359 U.S. 121, 124–39 (1959) (holding that defendant acquitted for robbery of a federally insured savings and loan association in federal court was not deprived of due process of law under the Fourteenth Amendment when he was subsequently tried in state court for same acts); Abbate v. United States, 359 U.S. 187, 195–96 (1959) (holding that conviction in state court for conspiracy to injure and destroy property of telephone companies in violation of state law did not bar subsequent federal prosecution for same acts under double jeopardy clause of the Fifth Amendment).

69. 16 U.S.C. §§ 1531–44 (2006).

70. *Id.* at § 1540(b).

71. *See, e.g.*, Lacey Act Amendments of 1981, 16 U.S.C. §§ 3371–78 (2006). *See generally* United States v. Bernal, 90 F.3d 465 (11th Cir. 1996) (affirming district court's holding that defendants violated various federal laws, including both the Lacey Act and the Endangered Species Act).

72. *See generally* MICHIGAN STATE UNIVERSITY COLLEGE OF LAW, ANIMAL LEGAL &

but approaching too close to nesting eagles or migrating whales is a crime as it causes them harm.[73] All states except Alaska, North Dakota, West Virginia, and Wyoming have enacted their own endangered species statutes.[74]

The possibility of prosecution under multiple federal and state statutes in endangered species cases raises double jeopardy concerns.[75] Therefore, endangered species cases can be compared to the prosecutions of those involved in the making of crush videos: the harm to the animal constitutes one offense, and the endangered species violation constitutes another offense.[76]

In an example of a prosecution involving violations of multiple federal laws, the government charged defendants with forty-two separate counts in *United States v. Hansen*.[77] The indictment read like a treatise on environmental law, as the defendants were charged with violating the Clean Water Act, the Comprehensive Environmental Response, Compensation, and Liability Act, the Resource Conservation and Recovery Act, and the Endangered Species Act and even faced a conspiracy allegation.[78] The Endangered Species Act was violated by "taking" an endangered species, a Wood Stork, through discharges of mercury into a marsh, creek, and river.[79] The defendants were convicted of most counts, demonstrating how a general intent to violate environmental statutes can result in multiple charges under a host of federal and state statutes.[80]

4. Animal Hoarders

Animal activists have noted that "[a]nimal hoarding is one of the greatest causes of animal suffering in the United States, and hoarders are responsible for causing more injuries, suffering, and deaths to animals than the intentionally cruel acts of violent animal abusers."[81] In animal hoarding cases, large numbers of animals, sometimes several hundred, are seized from a home.[82] Kept under extremely poor conditions, many animals die

---

HISTORICAL CENTER, http://www.animallaw.info/ (last visited Dec. 21, 2011).

73. *See id.*

74. *See id.*

75. *See infra* Part IV.

76. United States v. Hansen, 262 F.3d 1217, 1231 (11th Cir. 2001) (charging a defendant under multiple acts).

77. *Id.*

78. *Id.*

79. *Id.*

80. *Id.* at 1232. In California, for example, any cruelty against a separate specimen of an endangered species constitutes a separate offense. CAL. PENAL CODE § 597(e) (West 2010).

81. Lisa Avery, *From Helping to Hoarding to Hurting: When the Acts of "Good Samaritans" Become Felony Animal Cruelty*, 39 VAL. U. L. REV. 815, 818 (2005).

82. *See id.* (discussing animal hoarding).

from disease and starvation.[83] Mitigating circumstances are often present due to the mental illness of the owner.[84] Accordingly, the recidivism rate is nearly 100% in these cases.[85]

"A key shortcoming in the present laws is the necessity of filing criminal charges in order to get access to the animals."[86] Additionally, the animals must be held as evidence.[87] Therefore, major charging and sentencing issues arise due to the large number of animals involved and the need to remove all animals from the environment.[88] "Even once authorities seize hoarders' animals, the hoarders inevitably accumulate more."[89]

5. Violent Offenders

Family violence can take many forms, including emotional, sexual, child, and elder abuse.[90] As companion animals have made their way into our homes, many have also become exposed to domestic violence.[91] Violent family members often use companion animals to intimidate and control

---

83. *Id.* at 827.

84. *See* ANIMAL LAW WELFARE, *supra* note 11, at 282–83. According to Favre, a significant percentage of animal hoarders are ultimately institutionalized or placed under some type of protective care. *Id.* However, Satz argues that because hoarding is often seen as a product of mental illness, the legal focus in these cases "is shifted from animal well-being and the consequences of human behavior to the current and future well-being of the human engaging in the behavior." Ani B. Satz, *Animals as Vulnerable Subjects: Beyond Interest Convergence, Hierarchy, and Property*, 16 ANIMAL L. 65, 93 (2009).

85. The Hoarding of Animals Research Consortium, "a joint venture between professionals from Tufts University, the Massachusetts Society for the Prevention of Cruelty to Animals, Massachusetts General Hospital, and others . . . formed to investigate the problem of animal hoarding from an interdisciplinary perspective," reports that, without an ongoing system of support, the rate of recidivism among animal hoarders is almost one hundred percent. Avery, *supra* note 81, at 819, 834. In fact, an old adage says that "[a]nimal hoarders will pick up a stray cat on the way home from the courthouse." *Id.* at 834.

86. ANIMAL LAW WELFARE, *supra* note 11, at 283.

87. *Id.*

88. *See* HAW. REV. STAT. § 711-1109.6 (2010) (providing that a person commits the crime of animal hoarding if he intentionally, knowingly, or recklessly possesses and fails to provide necessary sustenance for more than fifteen dogs or cats). In a case from Connecticut, *State ex rel. Gregan v. Koczur*, 947 A.2d 282, 286 (Conn. 2008), officers found forty-six cats in a 950 square foot residence; a dead cat in the freezer; cat litter boxes filled with feces, vomit and urine; and moldy and insufficient cat food.

89. Avery, *supra* note 81, at 834. Illinois passed the nation's first animal hoarding law, prescribing felony criminal consequences and increases in penalties for subsequent offenses. *See* 510 ILL. COMP. STAT. ANN. 70/2.10 (2011).

90. *See generally* Caroline Forell, *Using* A Jury of Her Peers *to Teach About the Connection Between Domestic Violence and Animal Abuse*, 15 ANIMAL L. 53 (2008) (discussing the potential to use animal abuse as a way to emotionally abuse people who care for the animals).

91. *Id.*

their human victims, buying their victim's compliance by threatening, torturing, or even killing, a pet.[92] In a recent sample of several of the largest domestic violence shelters around the country, the Humane Society of the United States found that 91% of adult victims and 73% of children mention incidents of companion animal abuse when they enter violence shelters.[93]

In October 2010, New York's Suffolk County created the nation's first animal abuse registry, open to the public, requiring people convicted of cruelty to animals to register or face jail time and fines.[94] In Indiana and Oregon, a person who knowingly kills an animal with the intent to threaten or terrorize a household member commits domestic violence animal cruelty, a felony.[95] In Arkansas, a person who commits "aggravated cruelty" to a dog, cat, or horse in the presence of a child receives an enhanced, consecutive sentence of up to five years.[96] In Tennessee, if a convicted animal abuser resides in a household with minor children or elderly individuals, the court may send notification of the conviction to the appropriate protective agencies.[97]

Cases involving multiple abused human and animal victims present unique charging and sentencing issues. They raise the question of whether domestic violence offenders will be charged and sentenced one way for

---

92. Diana Wempen, *Four-Footed and Largely Forgotten: Exploring the Connections Between Animal Abuse and Domestic Violence*, ANIMAL ABUSE ISSUE (ABA Comm'n on Domestic Violence, Washington, D.C.), Summer 2007, at 1–2, *available at* http://www.abanet.org/domviol/docs/Wempen.pdf.

93. Humane Society of the United States, *Starting a Safe Havens for Animals Program* 1 (2004), *available at* http://www.humanesociety.org/assets/pdfs/2004_Safe Havens_Guide.pdf.

94. SUFFOLK COUNTY, N.Y., LOCAL LAW No. 55-2010, ch. 207, art. IV (2010). California Senate Bill 1277 would have created a database of persons over the age of 18 who have been convicted of animal abuse crimes. S.B. 1277, 2009–2010 Sess. (Cal. 2010), *available at* http://info.sen.ca.gov/pub/09-10/bill/sen/sb_1251-1300/sb_1277_bill_20100219_introduced.html. After passing the Senate Judiciary Committee in April 2010, S.B. 1277 "failed to advance further due to exorbitant cost estimates provided by the California Department of Justice." *Animal Abuse Registry Proposed in California*, ANIMAL LEGAL DEFENSE FUND (June 14, 2010), http://www.aldf.org/article.php?id=1274. In Tennessee, H.B. 1743 and S.B. 1590 would establish "registration, verification, and tracking requirements for an animal abuser . . . ." H.B. 1743, S.B. 1580, 107th Gen. Assemb., Reg. Sess. (Tenn. 2011), summary *available at* http://wapp.capitol.tn.gov/apps/billinfo/BillSummaryArchive.aspx?BillNumber=HB1743&ga=107. The requirements under the bill "are similar to the present law requirements for sex offenders." *Id.* "A violation would be a class E felony punishable only by fine." *Id.* The bill was assigned to the Agriculture Committee on February 24, 2011. *Bill Information for HB1743*, TENNESSEE GENERAL ASSEMBLY, http://wapp.capitol.tn.gov/apps/BillInfo/Default.aspx?BillNumber=HB1743 (last visited Dec. 21, 2011).

95. IND. CODE ANN. § 35-46-3-12.5 (LexisNexis 2010); OR. REV. STAT. § 167.320 (2010).

96. ARK. CODE ANN. § 5-4-702 (2010).

97. TENN. CODE ANN. § 39-14-212 (2010); *see also* D.C. § CODE 22-1002.01 (2010).

their human victims and another way for their animal victims. For instance, if an offender assaults his wife and child, he can be charged with two criminal counts, one for each human victim. But what if he also assaults his two dogs? Can he be charged with two counts of animal abuse? The answer depends upon the state in which the crime occurs.

6. Acceptable Animal Cruelty

Anti-cruelty statutes contain both definitions of offenses and lists of justifications. State animal cruelty laws provide specific exemptions for certain socially accepted practices (even those resulting in pain or death) such as scientific research, humane destruction of an animal, accepted veterinary practices, and lawful fishing, hunting, or trapping.[98] The excused acts are lawful and justified in spite of any harm they cause because the harm is thought to be outweighed by the benefits reaped from engaging in the conduct. Thus, the benefit is usually believed to overcome the harm even if the lawful act causes more harm than necessary to animals.[99]

Anti-cruelty statutes are intended to protect animals from the kinds of behavior that no responsible hunter or farmer would defend.[100] The ability to prosecute those who slowly kill and torture animals caught in the wild or who allow livestock to starve to death is significantly reduced in states that exempt abusers of wildlife or farm animals.[101]

About twenty states have enacted humane slaughter laws.[102] Nearly all of the statutes provide that an animal must be "rendered insensible to pain"

---

98. A number of states set forth the crime of animal cruelty and list potential defenses. *See, e.g.*, ALASKA STAT. §§ 11.61.140(a)–(c) (2010); IND. CODE ANN. § 35-46-3–12 (2010); MONT. CODE ANN. §§ 45-8-211(1)–(4) (2009); TEX. PENAL CODE ANN. §§ 42.092(b)–(d) (Vernon 2011).

99. For example, in a scientific experiment, there is rarely a governmental investigation into whether the harm done to the animals is greater than necessary. *See* Katherine M. Swanson, Note, *Carte Blanche for Cruelty: The Non-Enforcement of the Animal Welfare Act*, 35 U. MICH. J. L. REFORM 937, 938 (2002); Cass R. Sunstein, *The Rights of Animals*, 70 U. CHI. L. REV. 387, 392–95 (2003).

100. *See infra* note 103.

101. *See, e.g.*, KY. REV. STAT. ANN. § 525.130 (West 2010) (Hunting, fishing, trapping, food or commercial purposes, and sporting activities are exempt from animal cruelty laws.); WYO. STAT. ANN. § 6-3-203(m)(ii) (2011) (excepting industry accepted agricultural and livestock practices from prohibited conduct).

102. *See* ARIZ. REV. STAT. ANN. § 3-2016 (2010); CAL. FOOD & AGRIC. CODE § 19501 (2010); COLO. REV. STAT. § 35-33-203 (2010); CONN. GEN. STAT. § 22-272(a) (2010); FLA. STAT. ANN. § 828.22 (2010); 510 ILL. COMP. STAT. ANN. 75/0.01–75/8 (2010); IND. CODE § 15-17-1-1 (2010); IOWA CODE § 189A.18 (2010); KAN. STAT. ANN. § 47-1401 (2010); ME. REV. STAT. ANN. tit. 22 § 2521 (2002); MD. CODE ANN. AGRIC. § 4-123.1 (2010); MICH. COMP. LAWS ANN. § 287.551 (2010); MINN. STAT. ANN. § 31.59 (2010); N.H. REV. STAT. ANN. §427:33 (2010); N.J. STAT. ANN. § 4:22-16.1 (2010); OHIO REV. CODE ANN. § 945.01 (West 2010); OR. REV. STAT. § 603.065 (2010); PA. CONS. STAT. § 2362 (2010); R.I. GEN.

(i.e., made unconscious or killed) prior to being slaughtered.[103] However, the penalties for violating these statutes are relatively lenient, especially considering the prevalence of large-scale commercial slaughter operations in this country.[104] Florida by far has the strictest penalty, making each violation a criminal misdemeanor and a civil wrong subject to a maximum fine of $10,000 per day.[105] West Virginia has a penalty scheme that increases the punishment for subsequent violations, including the possible revocation of a license to slaughter.[106] In contrast, New Jersey has a provision explicitly exempting violators from penalty if the violation is considered "incidental" or "minor."[107]

In an example of the abuse suffered by large numbers of farm animals, in 2007, an undercover video shot by an investigator for the Humane Society of the United States showed downer cattle being abused by plant workers at the Hallmark and Westland meatpacking companies.[108] The San

---

LAWS ANN. § 4-1-2 (2011); VT. STAT. ANN. tit. 6 §3131 (2010); WASH. REV. CODE §§ 16.50.100-16.50.900 (2011); W. VA. CODE § 19-2E-1 (2010).

103. *See, e.g.*, CONN. GEN. STAT. § 22-272a (2010), ("No person engaged in business as a slaughterer, packer or stockyard operator shall cause or permit any cattle, calves, sheep, swine, horses, mules, goats or other animals to be slaughtered or put into position for slaughter unless such animals are rendered insensible to pain or are restrained by an approved method.").

104. *See, e.g.*, 510 ILL. COMP. STAT. ANN. 75/6 (2010) ("Any violation of this Act or of the rules and regulations promulgated by the Director is a petty offense."); N.J. STAT. ANN. § 4:22-16.1(b)(2) (2010) ("[N]o person may be cited or arrested for a first offense involving a minor or incidental violation . . . unless that person has first been issued a written warning.").

105. FLA. STAT. ANN. § 828.26 (2010).

106. *See* W. VA CODE ANN. § 19-2B-6 (2010).

107. N.J. STAT. ANN. § 4:22-16.1b(2) (West 2010). In 1996, the New Jersey state government directed its Board of Agriculture and Department of Agriculture to adopt humane standards for the raising of agricultural animals. *See* N.J. Soc. for Prevention of Cruelty to Animals v. N.J. Dept. of Agric., 955 A.2d 886, 888 (N.J. 2008). The New Jersey Society for the Prevention of Cruelty to Animals later sued the department alleging that various husbandry practices harm animals without providing a benefit to the animal, while only providing a slight convenience to the animal handler. *Id.* at 904–05. The New Jersey Supreme Court overturned the lower court's holding that agricultural institutions, rather than the State Board of Agriculture, may determine whether their own practices are humane. *Id.* at 905–07. But the court indicated that all but one of the husbandry practices at issue, tail docking, could be humanely performed. *Id.* at 908–09. *See Satz*, *supra* note 84, at 87–88, for a discussion of the routine husbandry practices allowed by the New Jersey Superior Court.

108. Andrenna Taylor, *From Downer Cattle to Mystery Meat: Chapter 194 Is California's Response to the Largest Beef Recall in History*, 40 MCGEORGE L. REV. 523, 523 (2009). A "downer" is a "[c]ommonly used term for animals that are disabled (nonambulatory) due to illness or injury." CONGRESSIONAL RESEARCH SERVICE, AGRICULTURE: A GLOSSARY OF TERMS, PROGRAMS, AND LAWS (2005), *available at* http://www.cnie.org/NLE/CRSreports/05jun/97-905.pdf.

Bernardino County, California District Attorney charged two workers with animal abuse, and both workers eventually pleaded guilty.[109]

In March 2009, in response to this tragedy and the ensuing public outcry, the USDA amended the federal meat inspection regulations to ban the slaughter of cattle that become non-ambulatory or disabled after passing initial inspection by Food Safety and Inspection program personnel.[110] Two months later, in May 2009, the United States Department of Justice announced it would seek to join the Humane Society in a federal lawsuit against the meatpacking companies for fraud and deception.[111] For the first time, the federal government allied itself with a non-profit animal protection organization in a lawsuit involving the mistreatment of farm animals.[112]

In another example of rapidly changing sensibilities, and perhaps in response to the Hallmark and Westland incident, a 2008 California initiative measure, "The Prevention of Farm Animal Cruelty Act," passed by a margin of 63% to 37%.[113] Beginning January 1, 2015, this measure prohibits the confinement of pregnant pigs, calves raised for veal, and egg-laying hens on a farm in a manner that does not allow them to turn around freely, lie down, stand up, and fully extend their limbs. Under this measure, any person who violates this law would be guilty of a misdemeanor, punishable by a fine of "up to $1,000 and/or imprisonment in county jail for up to six months."[114]

---

109. Will Bigham, *Westland/Hallmark Slaughterhouse Worker Sentenced to 270 Days in Jail for Cow Abuse*, SAN BERNARDINO COUNTY SUN (Sept. 25, 2008).

110. 9 C.F.R. § 309.3(e) (2011); *see* Petition to Amend 9 C.F.R. § 309.3(e) to Prohibit the Slaughter of Non-Ambulatory Pigs, Sheep, Goats, and Other Livestock and to Require that Such Animals be Humanely Euthanized, *available at* http://www.fsis.usda.gov/PDF/Petition_Humane_Handling.pdf.

111. Humane Society of the United States, *Meat Supplier Faces $150 Million Lawsuit for Using Sick, Injured Animals in School Lunch Program* (May 1, 2009), *available at* http://www.humanesociety.org/news/press_releases/2009/05/hallmark_false_claims_act_050109.html.

112. Megan A. Senatori & Pamela D. Frasch, *The Future of Animal Law: Moving Beyond Preaching to the Choir*, 60 J. LEGAL EDUC. 209, 228 (2010).

113. *California Passes Prevention of Farm Animal Cruelty Act (Proposition 2) by Ballet Initiative*, THE AGRICULTURAL LAW BRIEF (Agricultural Law Resource & Reference Center, Penn State Dickinson School of Law), Nov. 30, 2008, at 1, *available at* http://law.psu.edu/_file/aglaw/Nov_30_08.pdf.

114. STATE OF CALIFORNIA, VOTER INFORMATION GUIDE, 16–17 (2008) [hereinafter VOTER GUIDE], *available at* http://www.voterguide.sos.ca.gov/past/2008/general/titlesum/prop2-title-sum.htm (providing information about Proposition 2, an initiative entitled "Standards for Confining Farm Animals"). In December 2010, a Modesto, California egg farmer sued the state and the Humane Society of the United States, asking for a judge to interpret and clarify California's new law. P.J. Huffstutter, *Egg Farmer Sues Over Cruelty Law*, L.A. TIMES (Dec. 9, 2010), *available at* http://articles.latimes.com/print/2010/dec/09/business/la-fi-chicken-cage-lawsuit-20101209.

In particular, farm animals raise challenging issues in the area of charging and sentencing. Mistreatment of farm animals falls under multiple statutes.[115] Moreover, farm animals of the same type may not be readily distinguishable, causing difficulty in drafting complaints of sufficient clarity. Some prosecutors join all farm animals together in one count; some file each head of cattle as a separate count; and still others consider the failure to feed multiple cows and pigs as two units of prosecution—one for the cows and one for the pigs.[116]

As American society develops a heightened sensitivity to animal abuse and increasing media coverage accelerates its awareness of issues involving abused animals, governmental agencies are being pressured to modify their regulations.[117] While animal abuse legislation and regulations are being drafted at a rapid rate,[118] all of the latter categories of animal abuse raise important double jeopardy issues that have not been resolved and should be included in the debate.

III. ALTERNATIVES AND WHY RESOLUTION IS IMPORTANT

A. Statutory Alternatives: Summary of the Variations

In most jurisdictions, animals are still considered property.[119] Nevertheless, implicitly and explicitly, courts have struggled with and continue to bend and modify this concept.[120]

---

115. The Legislative Analysis for Proposition 2 states: "Other [California] laws specifically related to farm animals generally focus on the humane transportation and slaughter of these animals. Depending upon the specific violation, an individual could be found guilty of a misdemeanor or felony punishable by a fine, imprisonment, or both." VOTER GUIDE, *supra* note 114, at 17.

116. *Compare* Boushehry v. State, 648 N.E.2d 1174, 1180 (Ind. Ct. App. 1995) (two geese, two counts), *with* Amrein v. State, 836 P.2d 862, 865 (Wyo. 1992) (nine units of livestock, one count). In Texas, a defendant was convicted of one count of cruelty to animals for his actions involving approximately 300 head of cattle and two donkeys. Westfall v. State, 10 S.W.3d 85, 88 (Tex. Ct. App. 1999). The prosecution presented evidence that the cattle were malnourished. *Id.* The Court of Appeals held that a rational trier of fact could have found that Westfall intentionally and knowingly tortured livestock by failing to provide feed or supply care for his livestock causing them to starve and die. *Id.* at 92. Several states have arbitrarily excluded livestock from the definition of "animals" protected by their animal cruelty laws. *See, e.g.*, IOWA CODE ANN. § 717B.1(1)(a) (West 2003); UTAH CODE ANN. § 76-9-301(1)(b)(ii) (2003); Pamela D. Frasch et al., *State Animal Anti-Cruelty Statutes: An Overview*, 5 ANIMAL L. 69, 70 (1999), *available at* http://www.animallaw.info/articles/arus 5animall69.htm.

117. *See* Senatori, *supra* note 112, at 213–39, for a discussion of "The Four Levers of Social Justice." Senatori and Frasch argue that public anger and awareness are essential components of society's acceptance of animal rights. *Id.*

118. *See* Otto, *supra* note 1, at 132–33.

119. Favre, *Development*, *supra* note 26, at 7–8 (discussing the property status of

Current statutes are also confusing, sometimes resulting in court rulings that do not make sense. For example, in Vermont, a defendant found with malnourished animals was charged with three counts of animal cruelty, yet the deputies recovered nine dogs, five cats, and one goat.[121] One can speculate that one count applied to the dogs, one to the cats, and one to the goat, but a clear explanation is lacking.[122] In Kansas, thirty-four "very thin" horses were seized; yet, the defendant was charged with only one count of cruelty to animals.[123] In Florida, seventy-seven poodles were discovered without food, water, and sufficient air, but only two counts were filed against the defendant.[124] These are but a few of the cases revealing the necessity of greater statutory direction to prosecutors in animal cruelty cases. Such guidance would help prosecutors decide whether to charge defendants on the basis of the number of animals injured, the type of animal, the class of animal, the severity of injuries, or in some other way.

Table I, on pages 1142-43, lists states that have clarified their positions on charging and sentencing in criminal cases involving perpetrators who have injured multiple animals. Several types of statutory variations exist, as discussed in the following sections.

1. Animal Cruelty Statutes Authorizing One Count per Animal

The Animal Legal Defense Fund recommends that each act of animal abuse should constitute a separate offense.[125] The statutes of Alaska, Arkansas, Louisiana, Montana, and Wyoming follow this approach.[126]

---

animals in early American law); Kruse, *supra* note 26, at 1675. Some scholars have suggested that an adequate resolution to the problem of animal cruelty will come only with the abandonment of animals' status as property. Friend, *supra* note 26, at 201–02, 260 (discussing the concept of animals as property); Kruse, *supra* note 26, at 1675.

120. *See, e.g.*, Morgan v. Kroupa, 702 A.2d 630, 633 (Vt. 1997) ("[M]odern courts have recognized that pets generally do not fit neatly within traditional property principles . . . Instead, courts must fashion and apply rules that recognize their unique status . . . .").

121. State v. Eldredge, 910 A.2d 816, 817 (Vt. 2006).

122. VT. STAT. ANN. tit. 13 § 352(4) (2010) ("[A person commits the crime of cruelty to animals if the person] deprives an animal which a person owns, possesses or acts as an agent for, of adequate food, water, [and] shelter . . . ."). The statute does not authorize separate counts per animal.

123. State v. MacFarlane, 769 P.2d 682 (Kan. 1989) (In an unpublished decision, the Court of Appeals reversed the defendant's conviction on procedural grounds). In a similar case, *State v. Blom*, 45 S.W.3d 519, 520 (Mo. 2001), the defendant was charged with one count of misdemeanor animal abuse under Missouri Revised Statutes section 578.012 for having cattle at large on ten to fifteen occasions. Approximately thirty of the defendant's cattle escaped and entered a neighbor's fields, damaging her crops. *Id.*

124. State v. Wilson, 464 So. 2d 667, 668 (Fla. 1985). Similarly, in Idaho, only one misdemeanor count was filed against defendants who administered a poison to three foxes owned by another individual. State v. Farnsworth, 10 P.2d 295, 297 (Idaho 1932).

125. Animal Legal Defense Fund, *ALDF Model Animal Protection Laws; Offender*

2. The Statute is Silent

In many states, prosecutors charge defendants with one count per animal injured even though their statutes do not authorize them to do so. For example, in a Texas case, the prosecutor and the court treated injured animals like human victims, and the defendant was charged with one count for each animal injured, despite the state's statutory silence on this issue.[127] The defendant was convicted of ten counts of cruelty to dogs and cats after officers recovered 172 live animals.[128] The defendant claimed that because the complaint contained no description of the animals, she did not know which animal was involved in each count.[129] The court concluded that pictures made available by the prosecutor were "an adequate way, perhaps the only way, to give notice."[130]

Similarly, in a 2003 animal cruelty case in Tennessee, each of the 101 counts in an indictment specifically identified an individual animal and described its injuries.[131] The defendant was convicted of forty-seven counts of animal cruelty.[132] In addition to ratifying the prosecutor's charging decision, even though the statute was silent, the court referred to the "victimized animals."[133] Likewise, in a 1974 Connecticut case, the defendant was charged with twenty-eight counts of failing to provide and supply twenty-eight animals with wholesome air, food, and water.[134] The court reversed the conviction because the defendant lacked control of the

---

Registration & Community Notification, *available at* http://www.aldf.org/downloads/OffenderRegistryModelLaw.pdf.

126. *See* ALASKA STAT. 11.61.140(b) (2010); ARK. CODE ANN. § 5-62-103 (West 2010); LA. REV. STAT. ANN. § 14:102.1 (2004); MONT. CODE ANN. 45-8-211(C) (2010); WYO. STAT. ANN. § 6-3-203(k) (2010). The District of Columbia and Rhode Island have both enacted statutes providing that an animal abuser is punished for each offense against an animal, but neither state defines an "offense." D.C. CODE § 22-1001 (2010); R.I. GEN. LAWS § 4-1-2 (2010).

127. *See* Mills v. State, 802 S.W.2d 400, 402 (Tex. Ct. App. 1991) (charging defendant with ten counts). The officer identified photographs of each of the animals on which charges were based, and testimony described each animal. *Id.* at 405. As stated by the court, for the purposes of the case at hand, the relevant statutory provisions read: "(a) A person commits an offense if he intentionally or knowingly: . . . (2) fails unreasonably to provide necessary food, care, or shelter for an animal in his custody; [or] . . . (4) transports or confines an animal in a cruel manner; . . . ." *Id.* at 403 (citing TEX. PENAL CODE ANN. § 42.11 (Vernon 1974)).

128. *Id.* at 402.

129. *Id.* at 403.

130. *Id.* at 404.

131. State v. Webb, 103 S.W.3d 799, 804, 813–19 (Tenn. 2003).

132. *Id.* at 804.

133. *Id.* at 830–31, 836.

134. State v. Yorczyk, 356 A.2d 169, 169–71 (Conn. 1974).

animals.[135] In Nebraska, a prosecutor actually named the animals—Dee, Ace, and Moon—in the complaint, and the court ratified the convictions.[136]

3. A Separate Offense if the Violation Continues to Another Day

In Illinois, a second or subsequent violation of an animal abuse statute is a felony, and "every day" that a violation continues constitutes a separate offense.[137] The language of the statute is mandatory, requiring courts to impose the additional punishment.[138]

4. Some Animals Are More Important than Others

A majority of jurisdictions have expanded the scope of anti-cruelty statutes to encompass the protection of animals not generally considered to be of significant economic value.[139] Most states, however, distinguish between individuals who harm companion animals and those who injure non-domesticated animals.[140]

---

135. *Id.*

136. *See* State v. Ziemann, 705 N.W.2d 59, 62 (Neb. 2005).

137. 510 ILL. COMP. STAT. 70/3 (2010).

138. In Illinois, if the plain language of a statute or city ordinance mandates a minimum penalty for each day an offense is committed, a court is obligated to impose a multiplied sentence or fine, without implying any "exceptions, limitations, or conditions" to the language of the statute or ordinance in determining the proper punishment. *See* City of Chicago v. Elevated Props., L.L.C., 840 N.E.2d 677, 685 (Ill. Ct. App. 2005).

139. *See* ALA. CODE § 13A-11-14 (2010); ARIZ. REV. STAT. ANN. § 13-2910 (2011); CAL. PENAL. CODE § 597 (West 2010); COLO. REV. STAT. § 18-9-202 (2010); CONN. GEN. STAT. § 53-247 (2011); GA. CODE ANN. § 16-12-4 (2011); KAN. STAT. ANN. § 21-4310 (2010); LA. REV. STAT. ANN. § 14:102.1 (2011); ME. REV. STAT. ANN. tit. 17 § 1031 (2010); MINN. STAT. ANN. § 343.21 (2010); NEV. REV. STAT. § 574.100 (2011); N.H. REV. STAT. ANN. § 644:8 (2010); N.Y. AGRIC. & MKTS. LAW § 373 (McKinney 2011); N.C. GEN. STAT. § 14-360 (2010); OKLA. STAT. ANN. tit. 21, § 1685 (2011); 18 PA. CONS. STAT. ANN. § 5511 (West 2011); TENN. CODE ANN. § 39-14-212 (2010); TEX. PENAL CODE ANN. § 42.092 (West 2010); UTAH CODE ANN. § 76-9-301 (2010); VA. CODE ANN. § 3.2-6570 (2010); WYO. STAT. ANN. § 6-3-203 (2010).

140. *See* Rebecca J. Huss, *Valuing Man's and Woman's Best Friend: The Moral and Legal Status of Companion Animals*, 86 MARQ. L. REV. 47, 68–71 (2002) (describing the different subcategories of animals); *see also* CAL. PENAL CODE § 598b (2010); COLO. REV. STAT. § 35-42-109 (West 2010); DEL. CODE ANN. tit. 11 § 1325 (2010); IDAHO CODE ANN. § 25-3502 (2011); 510 ILL. COMP. STAT. ANN. 70/3.02 (2011); KAN. STAT. ANN. § 21-3727 (2010); LA. REV. STAT. ANN. § 14:102 (2011); MINN. STAT. ANN. § 343.01 (2011); MISS. CODE ANN. § 97-41-16 (2010); MONT. CODE ANN. § 45-8-211 (2009); NEB. REV. STAT. § 28-1008 (2010); NEV. REV. STAT. § 574.100 (2010); N.H. REV. STAT. ANN. § 644:8 (2010); N.J. STAT. ANN. § 4:22-26 (West 201); OHIO REV. CODE ANN. § 959.131 (West 2010); OR. REV. STAT. ANN. § 609.500 (West 2010); S.D. CODIFIED LAWS § 40-1-1 (2011); TENN. CODE ANN. § 39-14-202 (2010); VA. CODE ANN. § 3.2-6570 (2010).

In Alabama, for example, engaging in acts of cruelty in the first degree toward a dog or cat is a felony, whereas performing identical acts on horses, cows, rabbits, or any other animal is only a misdemeanor.[141] In Arkansas, each act of aggravated cruelty to a dog, cat, or horse is a felony, whereas the same act is a misdemeanor if another kind of animal is harmed.[142] In California, every intentional act against an endangered or threatened species constitutes a separate offense.[143]

5. One Count per Animal if a Certain Type of Abuse or Abuser

In Georgia, each violation of the animal protection statute regulating pet dealers, animal shelters, kennels, and stables constitutes a separate offense, even though Georgia's animal cruelty statute contains no such provision.[144] In New Jersey, a defendant is charged with a separate offense for each animal that is not euthanized as rapidly and as painlessly as possible.[145]

---

141. ALA. CODE §§ 13A-11-14, 13A-11-241 (2010).

142. Arkansas Code Annotated section 5-62-104 (2011) provides, in relevant part:

> (a) A person commits the offense of aggravated cruelty to a dog, cat, or horse if he or she knowingly tortures any dog, cat, or horse.
> (b) A person who pleads guilty or nolo contendere to or is found guilty of aggravated cruelty to a dog, cat, or horse:
> (1) Shall be guilty of a Class D felony[.]

ARK. CODE ANN. § 5-62-104 (2011).

However, Arkansas' general cruelty to animals statute merely prescribes that "[a]ny person who pleads guilty or *nolo contendere* to or is found guilty of cruelty to animals is guilty of an unclassified misdemeanor . . . ." ARK. CODE ANN. § 5-62-103(c) (2011). Only upon "a fourth or subsequent offense occurring within (5) five years of a previous offense of cruelty to animals" is a defendant guilty of a Class D felony. ARK. CODE ANN. § 5-62-103(f)(1) (2011).

143. CAL. PENAL CODE §§ 597(c)–(e) (West 2010).

144. *Compare* GA. CODE ANN. § 4-11-16 (2011) (For pet dealers or operators of animal shelters, kennels, and stables, "[e]ach violation of this article shall constitute a separate offense."), *with* GA. CODE ANN. § 16-12-4(b) (2011) ("A person commits the offense of cruelty to animals when he or she causes death or unjustifiable physical pain or suffering to any animal by an act, an omission, or willful neglect.").

145. N.J. STAT. ANN. § 4:22-19 (West 2011) (prohibiting killing an animal "by hypoxia induced by decompression or in any other manner, by the administration of a lethal gas other than an inhalant anesthetic, or in any other manner except by a method of euthanasia generally accepted by the veterinary medical profession as being reliable, appropriate to the type of animal upon which it is to be employed, and capable of producing loss of consciousness and death as rapidly and painlessly as possible"). The statute provides that "[e]ach animal destroyed . . . shall constitute a separate offense." *Id.*

6. One Count per Animal in Oregon?

Oregon has taken a unique approach to the problem of multiple violations of the same statutory provision arising from a single criminal episode. This approach, however, may not be of much help in animal abuse cases. Where the violations involve multiple victims, Oregon's general rule is that there are as many violations as there are victims.[146] However, when the case involves property crimes such as theft or criminal mischief, there is only one offense.[147] Although the statutory provision appears in the same title as Oregon's animal abuse statutes, it is unclear whether it applies to animals.[148] Nevertheless, at least one Oregon court found that an animal abuse conviction was similar to the "person" crime of first-degree assault because "animal abuse, like assault, involves the abuse of a living thing that can feel pain."[149] In addition, the term "assault," which is usually associated with crimes against humans, is used to define "sexual assault of an animal" and "assault of a law enforcement animal."[150]

Forty states and the District of Columbia have animal cruelty statutes that are completely silent on the issue of multiple animals injured at the same time.[151] States not listed in Table I have statutes that are silent on the issue. Table I lists solutions developed by the remaining jurisdictions:[152]

---

146. *See* OR. REV. STAT. § 161.067 (2009).

147. *Id.*

148. Chapter 167 regulates offenses against public health, decency, and animals.

149. State v. Agee, 196 P.3d 1060, 1068 (Or. 2008).

150. OR. REV. STAT. §§ 167.333, 167.339 (2009).

151. Alabama, Arizona, Colorado, Connecticut, Delaware, the District of Columbia, Florida, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Maine, Maryland, Massachusetts, Minnesota, Mississippi, Missouri, Nebraska, Nevada, New Hampshire, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, and Wisconsin.

152. Illinois and Oregon are included in Table I because, though neither has an animal cruelty statute speaking directly to the issue of multiple animals injured at the same time, each has a statute that is relevant to the discussion of criminal penalties for animal abuse.

TABLE I

ANIMAL CRUELTY STATUTES

One Count Per Animal Always	**Alaska** ALASKA STAT. § 11.61.140(b) (2010) ("Each animal that is subject to cruelty to animals . . . shall constitute a separate offense."). **Wyoming** WYO. STAT. ANN. § 6-3-203(k) (2011) ("Each animal affected by the defendant's conduct may constitute a separate count . . . under this section."). **Louisiana** LA. REV. STAT. ANN. § 14:102.1(A)(3) (2011) ("[I]f more than one animal is subject to an act of cruel treatment . . . , each act shall constitute a separate offense."); LA. REV. STAT. ANN. § 14:102.1(B)(7) (2011) ("[W]here more than one animal is tortured . . . or where more than one head of livestock is tampered with, each act comprises a separate offense."). **Montana** MONT. CODE ANN. § 45-8-211(2)(c) (2009) ("[W]hen more than one animal is subject to cruelty to animals, each act may comprise a separate offense."). **Arkansas** ARK. CODE ANN. § 5-62-103 (West 2010) (Each act of cruelty to animals committed against more than one animal may constitute a separate offense.); ARK. CODE ANN. § 5-62-104(e) (West 2010) ("[E]ach alleged act of the offense of aggravated cruelty to a dog, cat, or horse committed against more than one (1) dog, cat, or horse may constitute a separate offense.").
One Count Per Animal if a Certain Kind of Animal	**California** CAL. PENAL CODE § 597(e) (West 2010) ("[E]ach act of malicious and intentional maiming, mutilating, or torturing a separate specimen of [endangered species or fully protected bird, mammal, reptile, or fish] is a separate offense."). **Pennsylvania** 18 PA. CONS. STAT. ANN. § 5511(e.1) (West 2011) ("A person who violates this subsection on a second or subsequent occasion commits a misdemeanor of the third degree for each equine animal transported.").

One Count Per Day the Violation Continues	**Illinois** 510 ILL. COMP. STAT. 70/17 (West 2011) ("Any person convicted of any act of abuse or neglect or of violating any other provision of this Act . . . is guilty of a Class B misdemeanor. A second or subsequent violation is a Class 4 felony with every day that a violation continues constituting a separate offense."); 70/3 (owner shall provide food, water, etc); 70/4 (selling dyed animals); 70/7 (confinement or detention during transport); 70/7.5(e) (downed animals).
One Count Per Animal if a Certain Type of Abuse or Abuser	**California** CAL. PENAL CODE § 597.7(b) (West 2010) ("[A] first conviction for violation of this section is punishable by a fine not exceeding one hundred dollars ($100) per animal."); CAL. PENAL CODE § 597o(b)(1) (West 2010) ("Any person who violates this section is guilty of a misdemeanor and is subject to a fine of one hundred dollars ($100) per equine being transported."). **Georgia** GA. CODE ANN. § 4-11-16 (2011) (For pet dealers or operators of animal shelters, kennels, and stables, "[e]ach violation of this article shall constitute a separate offense."). **Michigan** MICH. COMP. LAWS § 750.50b(3)(b) (West 2011) (imposing "[a] fine of not more than $5,000.00 for a single animal and $2,500.00 for each additional animal involved in the violation"). **New Jersey** N.J. STAT. ANN. § 4:22-19, 4:22-19.4 (2010) ("Each animal destroyed [in an inhumane manner] shall constitute a separate offense.").
One Count Per Animal in Oregon?	**Oregon** O.R.S. § 161.067 (West 2008) ("(1) When the same conduct or criminal episode violates two or more statutory provisions and each provision requires proof of an element that the others do not, there are as many separately punishable offenses as there are separate statutory violations. (2) When the same conduct or criminal episode, though violating only one statutory provision involves two or more victims, there are as many separately punishable offenses as there are victims.").

B. Why Resolution of This Disparity and Ambiguity is Important

In cases involving multiple animal victims, clear charging and sentencing statutes are critical in order to eliminate disparity and provide guidance to prosecutors. Without these guidelines, prosecutors will (1) sometimes file multiple counts using each instance of cruelty as a distinct offense;[153] (2) occasionally elevate the level of the offense;[154] (3) frequently file one count for all injured animals;[155] (4) sometimes separate animals by species;[156] and (5) sometimes base charges only on the most egregious and easily proven instances of cruelty.[157]

One of the basic premises of Anglo-American criminal law, based on the notion of fair play, is that the public must receive some advance warning outlining what criminal conduct is and how such conduct will be punished.[158] "No one may be required at peril of life, liberty or property to speculate as to the meaning of penal statutes. All are entitled to be informed as to what the State commands or forbids."[159] When a jurisdiction fails to provide a clear governing statute, defendants are without notice that their conduct is criminal or are unaware what consequences their actions will bring. Therefore, unclear statutes raise serious constitutional issues concerning the rights of criminal defendants that are difficult for courts to treat.[160]

The broad purposes of criminal law are to encourage desirable conduct and to inhibit behavior that society considers harmful or undesirable.[161]

---

153. State v. Yorczyk, 356 A.2d 169, 170 (Conn. 1974); State v. Webb, 103 S.W.3d 799, 803 (Tenn. Crim. App. 2003); Mills v. State, 802 S.W.2d 400, 405–06 (Tex. Crim. App. 1991).

154. United States v. Bernal, 90 F.3d 465, 466 (11th Cir. 1996); United States v. Atkinson, 966 F.2d 1270, 1275–76 (9th Cir. 1992).

155. Amrein v. State, 836 P.2d 862, 865 (Wyo. 1992). *See generally* Westfall v. State, 10 S.W.3d 85 (Tex. Ct. App. 1999) (Prosecutor filed charges against defendant for harm caused to multiple animals.).

156. *See generally* State v. Eldredge, 910 A.2d 816 (Vt. 2006) (different charges for different species of animals).

157. *See generally* State v. Webb, 103 S.W.3d 799 (Tenn. Crim. App. 2003) (Prosecution proved charges for outrageous and egregious conduct.).

158. *See* United States v. Apollo Energies, Inc., 611 F.3d 679, 681 (10th Cir. 2010). "First, due process requires citizens be given fair notice of what conduct is criminal. A criminal statute cannot be so vague that 'ordinary people' are uncertain of its meaning." *Id.* (citing Kolender v. Lawson, 461 U.S. 352, 357 (1983)); *see also* LAFAVE, *supra* note 68, at 10.

159. *Apollo Energies, Inc.*, 611 F.3d at 681 (quoting Lanzetta v. New Jersey, 306 U.S. 451, 453 (1939)).

160. *See generally* PAUL H. ROBINSON, CRIMINAL LAW DEFENSES, MULTIPLE OFFENSE LIMITATIONS, § 68(d) (1984) (discussing the Model Penal Code and multiple convictions for offenses that essentially prohibit the same conduct).

161. LAFAVE, *supra* note 68, at 30.

Criminal punishment aims to deter criminals from committing further crimes by giving them an unpleasant experience that they will not want to repeat.[162] Charging and sentencing standards are critical components of prevention. The more offenses for which a defendant may be charged and the greater the consequences of each offense, the more significant the deterrent effect will be.[163]

Statutes specifically defining multiple offenses and clarifying what conduct is and is not appropriate help to change societal perceptions about the value of animals.[164] To raise awareness of animal abuse and to prevent its continued existence, many states are enhancing the sentences of repeat animal abusers. For example, Illinois counts every day that a violation continues as a separate offense;[165] California[166] and Pennsylvania[167] enhance the sentences of those who abuse certain kinds of animals; California,[168] Georgia,[169] Michigan,[170] and New Jersey[171] punish certain select abusers and abuses in a greater way, and Michigan[172] increases the sentences of those who harm multiple animals. Other states have passed statutes requiring prosecutors to charge defendants for each abused animal as a separate offense.[173] All of these policies give animals more respectability under the law.

IV. DOUBLE JEOPARDY AND THE UNIT OF PROSECUTION

When the same statutory violation is charged twice, the double jeopardy clause of the Fifth Amendment requires courts to determine whether the facts underlying each count were intended to constitute separate "units" of prosecution.[174] A unit of prosecution is the manner in which a criminal statute permits a defendant's conduct to be divided into discrete acts for prosecuting multiple offenses by establishing whether the conduct consists of one or more violations of a single statutory provision.[175]

A defendant may not be convicted of more than one offense if the offense is defined as a continuing course of conduct, and his or her course

---

162. *Id.* at 31.

163. *Id.* at 780.

164. *Id.*

165. 510 ILL. COMP. STAT. ANN. 70/17 (West 2011).

166. CAL. PENAL CODE § 597(e) (West 2010).

167. 18 PA. CONS. STAT. ANN. § 5511(e.1) (West 2010).

168. CAL. PENAL CODE § 597.7(b) (West 2010).

169. GA. CODE ANN. § 4-11-16 (2011).

170. MICH. COMP. LAWS ANN. § 750.50b(3)(b) (West 2011).

171. N.J. STAT. ANN. § 4:22-19 (West 2011).

172. *See generally* MICH. COMP. LAWS. ANN. § 750.50b; Otto, *supra* note 1, at 139–40 (discussing Michigan's statutory changes increasing penalties for offenders).

173. *See supra* Table I.

174. United States v. Ansaldi, 372 F.3d 118 (2d Cir. 2004).

175. *See* Robinson, *supra* note 153, at § 68 (discussing multiple offenses).

of conduct was not interrupted, unless the law provides that specific periods of such conduct constitute separate offenses.[176]

A foundational principal in determining the "unit of prosecution" for any crime is to look at the goals of the law.[177] Criminal laws define various interests that society seeks to protect such as protection of the person from physical harm; protection of property from loss, destruction, or damage; protection of reputation from injury; protection against sexual immorality; protection of the government from injury or destruction; protection against interference with the administration of justice; protection of public health; and protection of public peace and order.[178]

Drawing on these common aims, the following paragraphs trace out the proper unit of prosecution in animal abuse cases. Part V proposes that each injured animal constitutes a legally recognizable "victim" that should form the basis of one count.

A. Ascertaining Units of Prosecution

We have already observed that the goal of modern animal abuse legislation is the protection of animals.[179] Accordingly, instead of looking at the wrong committed by the perpetrator of animal abuse, we should focus on the harm done to the animals. In order to achieve the purpose of animal abuse legislation, should each harmed animal constitute a unit of prosecution? If so, then an examination of human victim charging and sentencing principles may provide useful guidance.

In cases involving human victims, the general rule is that the number of people directly harmed or threatened during the commission of a crime can constitute a unit of prosecution.[180] For example, when a defendant is charged with the homicide of different victims and is found guilty on each count, he may be sentenced separately on each count and these sentences can run consecutively.[181] This is true because the killing of different persons constitutes separate crimes "even though done at the same time with one stroke of the same death-dealing instrument[:] while the stroke was one transaction, the killing of different persons with that stroke constitutes several criminal transactions."[182] Thus, in crimes against the person, as

---

176. MODEL PENAL CODE § 1.07 (1962).

177. LAFAVE, *supra* note 68, at § 1.5.

178. *Id.*

179. *See supra* Part II.

180. State v. Myers, 298 S.E.2d 813, 815 (W. Va. 1982); Robinson, *supra* note 153, at § 68.

181. Brown v. State, 201 S.E.2d 14, 16 (Ga. Ct. App. 1973).

182. *Id.* at 16. In *State v. Myers*, 298 S.E.2d 813, 814–15 (W. Va. 1982), the defendant collided with a car going in the opposite direction, killing two people. He was found guilty of two counts of involuntary manslaughter, and his sentences were directed to be served consecutively. *Id.* at 814–15. The court held that when charged under either the negligent

contrasted with crimes against property, a single criminal act or episode generally may constitute as many offenses as there are victims.[183]

The general principle in property cases is that a "single crime cannot be fragmented into more than one offense."[184] For example, in California, a couple convicted of nine counts of possessing property with altered serial numbers, including four television sets, two pairs of wood speakers, a stereo component system, a tape deck, and a clock radio, was found to have possessed only one classification or count of contraband.[185] In another California case, a defendant charged with and convicted of eleven counts of possessing eleven identical blank checks on the same date, was found to have committed one crime, not eleven.[186]

Two United States Supreme Court decisions, however, illustrate how a multiple-count conviction can be appropriate in a criminal case involving

---

homicide statute or the involuntary manslaughter statute for multiple deaths resulting from a single act, a defendant may receive as many consecutive sentences as there were deaths. *Id.* at 815. The rationale behind this rule is that when a "crime is committed against people rather than property, the general rule is that there are as many offenses as there are individuals affected." *Id.*

183. The appropriate unit of prosecution for crimes of violence is the person assaulted or killed, not the underlying criminal act. *See* Commonwealth v. Melton, 763 N.E.2d 1092, 1096 (Mass. 2002); 40A AM. JUR. 2D *Homicide* § 181 (2008).

In *Melton*, the defendant was convicted of, among other things, four counts of assault with a dangerous weapon. *Melton*, 763 N.E.2d at 1094. As the defendant only fired a single shot into a vehicle occupied by four people, he argued that he could not have intended to commit more than a single battery as it would have been physically impossible to hit all four victims with a single shot. *Id.* at 1094–96.

The Supreme Judicial Court of Massachusetts affirmed and held that the evidence supported defendant's transferred intent to assault all four victims because "once [the mens rea is] established as to any victim, it satisfies that element with respect to all other victims, even if those victims are unintended or even unknown to the defendant." *Id.* at 1098. It does not matter whether or not unintended victims were actually shot. The victims' injuries only affect whether the crime is an assault or battery. *Id.* at 1098. In response to defendant's impossibility argument, the Court noted that "while the laws of physics may determine the extent of injury or physical damage accomplished by a person's act, the perpetrator's intent is not necessarily so constrained. A person can intend things that are hopelessly unrealistic or even absurd." *Id.* at 1097.

This is not to say that there are never multiplicity issues in cases involving one act and multiple possible human victims, such as with statutes related to feticide, drunk driving, and assault. However, a complete survey of those statutes is beyond the scope of this article.

184. People v. Rouser, 69 Cal. Rptr. 2d 563, 568 (Cal. Ct. App. 1997).

185. People v. Harris, 139 Cal. Rptr. 778, 786 (Cal. Ct. App. 1977).

186. People v. Bowie, 140 Cal. Rptr. 49, 58 (Cal. Ct. App. 1977). In a West Virginia case in which a defendant took and cashed a number of checks made payable to a hospital, the court held that "if different evidence is required to prove each count in an indictment, then each count states a separate offense and a conviction will support a separate sentence." State v. Shafer, 284 S.E.2d 916, 920 (W. Va. 1981) (citing United States v. Hale, 468 F.2d 435 (5th Cir. 1972)).

multiple items of property. In *Blockburger v. United States,* the Court held that each of several successive sales made to the same person constituted a distinct offense, however closely they followed each other.[187] The Court reasoned that "[t]he first transaction, resulting in a sale, had come to an end. The next sale was not the result of the original impulse, but of a fresh one—that is to say, of a new bargain."[188]

In *Ebeling v. Morgan*, a defendant willfully opened seven mail bags and was charged with seven separate counts of "injury to mail bags."[189] Each count specifically described an individual mail pouch.[190] The Court found that the words of the statute plainly indicated that Congress intended to protect each mail bag from injury and mutilation.[191] Although the transaction of cutting the mail bags was in a sense continuous, the complete statutory offense was committed every time a mail bag was cut in the manner described.[192] According to the Court, the crime committed in *Ebeling* was not a continuous offense, where the crime is necessarily a single one, though committed over a period of time.[193] In support of this finding, the Court contrasted *Ebeling* with a prior case, *In re Snow*, in which it found "an attempt . . . to divide into separate periods of time the offense of continuous cohabitation with more than one woman" an improper attempt to cut a continuous offense into separate crimes.[194] Each of the sales in *Blockburger* constituted a separate offense because each one had ended before the next one began.[195] Each and every mailbag in *Ebeling* constituted a separate offense because the statute clearly specified Congress' intention.[196]

Thus, under the view that animals are analogous to human victims, a single criminal act or episode could constitute as many offenses as there are "victims." Similarly, if animals are analogized to property, each injured animal could constitute a separate offense only if each act of abuse is a new and complete offense separated in time from the prior offense. If the animal abuse is continuing, only one offense is committed. Separate acts of animal abuse, however, would always constitute a separate offense if clearly authorized by statute.

---

187. Blockburger v. United States, 284 U.S. 299, 304 (1932).
188. *Id.* at 303.
189. Ebeling v. Morgan, 237 U.S. 625, 627 (1915).
190. *Id.* at 627–28.
191. *Id.* at 629.
192. *Id.*
193. *Id.* at 629–30.
194. *Id.* at 630 (citing *In re* Snow, 120 U.S. 274 (1887)).
195. *See* Blockburger v. United States, 284 U.S. 299, 303 (1932).
196. *See Ebeling*, 237 U.S. 625, 627–28 (1915).

B. The Struggle in Defining the Unit of Prosecution in Animal Abuse Cases

Cases from Wyoming,[197] Ohio,[198] and California[199] illustrate the struggle courts experience with the issue of whether animals should be compared to human victims, property, or some new category of sentient beings. Due to the lack of statutory guidance in most states, many courts must resolve the vagueness in their own fashion. In Wyoming, statutory vagueness resulted in the court's reversal of multiple punishments;[200] in Ohio, vagueness was interpreted to favor the abused animals;[201] while in California, animals were analogized to victimized children.[202]

In Wyoming, in *Amrein v. State*, forty horses and forty-five cows had been deprived of food, water, and shelter, which led to the defendant being charged with nine counts of animal cruelty—six counts referred to horses and three to cows—and sentenced to eight consecutive jail terms.[203] In comparing animals to narcotics, the court found that the defendant's double jeopardy rights were violated because the multiple counts of cruelty to animals resulted from a single continuous transaction.[204]

In this case, the prosecution argued that the Wyoming legislature intended to protect individual animals, which individually feel pain and suffering.[205] The defense, however, maintained that the statutory provision was susceptible to an interpretation that the "unit of prosecution" should be determined by the defendant's conduct of failing to provide for "the animals."[206] The court held that there was one transaction—one common scheme or plan—and that because one offense was merged with another, the convictions could not be stacked.[207] Therefore, multiple consecutive sentences would constitute multiple punishments for the same offense at one trial.[208] Accordingly, in Wyoming, rules interpreting property offenses apply to animals, and animals, even of different species, are joined together.

Two Ohio cases illustrate an opposing view. The first, *Ohio v. Lapping*, involved a defendant's failure to provide adequate food and water to a herd of twenty-eight cows.[209] There, the trial court declined a request to treat the twenty-eight counts as allied offenses and to merge the counts into one for

197. Amrein v. State, 836 P.2d 862 (Wyo. 1992).

198. State v. Angus, No. 05AP-1054, 2006 WL 2474512 (Ohio Ct. App. Aug. 29, 2006); State v. Lapping, 599 N.E.2d 416 (Ohio Ct. App. 1991).

199. People v. Sanchez, 114 Cal. Rptr. 2d 437, 439–40 (Cal. Ct. App. 2001).

200. *Amrein*, 836 P.2d at 865.

201. Ohio v. Lapping, 599 N.E.2d 416, 422–23 (Ohio Ct. App. 1991).

202. *Sanchez*, 114 Cal. Rptr. 2d at 444.

203. *Amrein*, 836 P.2d at 863.

204. *Id.* at 865.

205. *Id.*

206. *Id.*

207. *Id.*

208. *Id.*

209. Ohio v. Lapping, 599 N.E.2d 416, 417 (Ohio Ct. App. 1991).

sentencing.[210] The appellate court concurred and applied the Ohio statute providing that "where the defendant's conduct constitutes two or more offenses of dissimilar import, or where his conduct results in two or more offenses of the same or similar kind committed separately or with a separate animus as to each, the indictment or information may contain counts for all such offenses, and the defendant may be convicted of all of them."[211] Analogizing the abused animals to victims of vehicular homicide, the court concluded that the defendant's conduct constituted two or more offenses based on dissimilar import and that each offense was committed separately or with separate animus.[212]

A later Ohio case, *State v. Angus*, extended the comparison of animals to human victims.[213] Convicted of negligently depriving two dogs of sustenance, the defendant claimed that the trial court erred in imposing consecutive terms.[214] The court concluded that where a defendant's conduct involves multiple "victims," the defendant should be convicted and sentenced for each offense.[215] Consequently, in Ohio, human victim principles are applied to animals.

In the California case of *People v. Sanchez*, the defendant contended that six of his seven convictions should have been reversed because the trial court failed to instruct the members of the jury that they had to agree unanimously on the particular animals that had been abused.[216] The

---

210. *Id.* at 423.

211. *Id.* at 422–23 (citing OHIO REV. CODE ANN. § 2941.25) (emphasis added). "Animus" is defined as "[i]ll will; animosity." BLACK'S LAW DICTIONARY 97 (8th ed. 2004).

212. *Lapping*, 599 N.E.2d at 422–23.

213. State v. Angus, No. 05AP-1054, 2006 WL 2474512, at *7 (Ohio Ct. App. Aug. 29, 2006).

214. *Id.*

215. *Id.*

216. People v. Sanchez, 114 Cal. Rptr. 2d 437, 439–40 (Cal. Ct. App. 2001). When a defendant is charged with a single criminal act, and the evidence reveals more than one such act about which the jury might disagree, a unanimity instruction is required. CAL. CONST. art. I, § 16; People v. Davis, 115 P.3d 417, 452 (Cal. 2005). "The requirement of unanimity as to the criminal act is intended to eliminate the danger the defendant will be convicted even though there is no single offense all the jurors agree the defendant committed." People v. Zavala, 30 Cal. Rptr. 3d 398, 405 (Cal. Ct. App. 2005).

Therefore, what is required is that the jurors unanimously agree that the defendant is criminally responsible for "*one discrete criminal event*." People v. Davis, 10 Cal. Rptr. 2d 381, 389 (Cal. Ct. App. 1992). In *Davis*, the Court explained that the defendant was entitled to a unanimity instruction where "evidence disclosed two distinct takings [(]the taking of Harris's car from Boyd and Harris, and the taking of Boyd's rings from her person[)]" and "the prosecutor argued that the jury could rely on either theory to convict defendant of the robbery." 115 P.3d at 452.

"The unanimity instruction is not required where the criminal acts are so closely connected that they form a single transaction or where the offense itself consists of a continuous course of conduct." People v. Rae, 125 Cal. Rptr. 2d 312, 317 (Cal. Ct. App. 2002). "The 'continuous conduct' rule applies when the defendant offers essentially the

defendant had been charged with the abuse of six to eight rabbits; several ducks; multiple chickens; more than twelve geese; the failure to give sustenance, shelter, and/or drink to dogs between 1997 and 1999; and the failure to provide medical treatment to a severely wounded puppy.[217] In upholding all but one of the defendant's convictions, the Court of Appeal found that the California animal abuse statute could be violated by subjecting an animal to needless suffering,[218] by inflicting unnecessary cruelty,[219] or by failing to provide an animal with food, drink, and shelter. The court ultimately concluded that the crime of animal cruelty could be committed either by a distinct act or by a continuous course of conduct.[220]

The first two theories outlined by the *Sanchez* court—subjecting an animal to needless suffering or inflicting unnecessary cruelty—could result from either a distinct act or a continuous course of conduct in the same way that child abuse may be committed by repetitive acts.[221] The failure to provide food, shelter, or water is a continuous course of conduct offense analogous to failing to provide these necessities to a child.[222]

The *Sanchez* court held that the defendant's failure to provide care for the severely wounded puppy "was a form of continuing neglect."[223] In a separate charge, the prosecutor introduced evidence of two distinct and unrelated incidents in which the defendant kicked a dog.[224] The court reversed the trial court on this count because the evidence established two discrete criminal events of cruelty to dogs, each sufficient to support a conviction and separately punishable and each subject to a different possible defense.[225]

The remaining five counts of animal abuse in *Sanchez*—pertaining to the rabbits, ducks, chickens, and geese—were based upon evidence establishing that the defendant failed to provide adequate food and water on an ongoing basis. By causing malnourishment and dehydration, these actions resulted in the needless suffering of animals and the death of many of them.[226] Because the Court found these charges to be based on a

---

same defense to each of the acts, and there is no reasonable basis for the jury to distinguish between them." People v. Stankewitz, 793 P.2d 23, 41 (Cal. 1990) (citation omitted). A continuing course of conduct has been held to exist where the wrongful acts were successive, compounding, and interrelated. *Sanchez*, 114 Cal. Rptr. 2d. at 444 (Cal. Ct. App. 2001); People v. Dieguez, 107 Cal. Rptr. 2d 160, 166 (Cal. Ct. App. 2001).

217. *Sanchez*, 114 Cal. Rptr. 2d. at 446.

218. *Id.* at 445.

219. *Id.*

220. *Id.*

221. *Id.* (citing People v. Avina, 18 Cal. Rptr. 2d. 511 (Cal. Ct. App. 2001)).

222. *Id.* (citing CAL. PENAL CODE § 270 (West 2008); People v. Morrison, 202 P.3d 348 (Cal. Dist. Ct. App. 1921)).

223. *Id.* at 446.

224. *Id.*

225. *Id.*

226. *Id.*

continuous course of conduct, no unanimity instruction was required.[227] Unlike the conviction based on the kicking of a dog, in support of which the prosecution presented evidence of two distinct and unrelated incidents in which the defendant kicked a dog, for these continuing course of conduct counts the jurors did not have to agree which particular animal constituted the basis for the conviction.[228] All of the animals were consolidated together, and the continuous course of conduct exception applied to exempt the prosecution from showing an exact date the animals were abused.[229]

Sanchez illustrates the need for charging guidelines in animal abuse cases. Not only did the indictment fail to specify the dog the state was referring to, but it also combined within its counts many of the animals by species, even though one count addressed the injuries to only one puppy.[230] *Sanchez* is also significant because the court analogized animal abuse to continuous offenses involving human victims such as the failure to provide for a minor child, annoying or molesting a child, child abuse, and contributing to the delinquency of a minor.[231]

In *People v. Counts*, an unpublished case decided by the California Court of Appeal, a detective found seventeen injured adult pit bulls and four puppies in the backyard of a Los Angeles residence.[232] A jury convicted Counts of "10 counts of unlawfully owning, possessing, keeping, and training a dog with the intent that the dog engage in an exhibition of fighting with another dog."[233] On appeal, Counts contended that "the trial court erred in failing to instruct the jury sua sponte regarding unanimity, or alternatively appellant should have been charged with only one count of [unlawful activity]."[234] The Court agreed, stating:

> Here, the prosecutor presented 10 identical counts and introduced evidence regarding 17 adult dogs that were being trained for, and had injuries consistent with, a staged dogfight. But, the jury was not instructed which of the 10 counts applied to which of the 17 dogs or instructed on the need to unanimously agree as to a specific dog as a basis for a specific count. Nor, in closing argument did the prosecutor specify which counts applied to which dogs. Accordingly, the People concede, and we agree,

---

227. *Id.*

228. *Id.*

229. *Id.* at 444 (citations omitted).

230. *Id.* at 446.

231. *Id.* at 444 (citations omitted). *Sanchez* was the first case to conclude that animal abuse is a continuing offense. *Id.*

232. People v. Counts, No. BA319187, 2010 WL 4869754 at *1 (Cal.Ct. App. Dec. 1, 2010).

233. *Id.*

234. *Id.*

that the trial court erred in failing to give a unanimity instruction and that counts 2 through 10 should be reversed.[235]

C. Michigan's Model

The Michigan Legislature has designed three primary provisions related to cruelty to animals: prohibitions against (1) negligent care; (2) intentional infliction of pain and suffering; and (3) participation in animal fighting.[236]

For neglect, if a violation involves one animal, the perpetrator is guilty of a misdemeanor punishable by imprisonment, a fine, and community service.[237] If a violation involves two or three animals, or the death of any animal, the perpetrator is punished by a longer term of imprisonment, a larger fine, and longer community service hours.[238] If the violation involves four to nine animals, or the perpetrator has a prior conviction, he or she is guilty of a felony punishable by imprisonment for not more than two years, a larger fine, and an increased number of community service hours.[239] If the violation involves ten or more animals, or the person had two or more prior convictions, the person is guilty of a felony punishable by imprisonment for not more than four years, a fine of not more than $5,000, and community service for not more than 500 hours.[240] In addition to these penalties, a perpetrator can also be convicted of any other violation of law arising out of the same transaction, and the court can order a term of imprisonment to be served consecutively.[241]

For intentional acts in Michigan, there is no additional punishment for prior convictions, no additional incarceration for multiple animals, no language punishing perpetrators for other offenses arising out of the same transaction, and no provision for consecutive sentences.[242] The statute provides that whoever knowingly kills, tortures, poisons, or mutilates *an* animal is guilty of a felony punishable by imprisonment for not more than four years, a fine of not more than $5,000 for a single animal and $2,500 for each additional animal involved in the violation, but not to exceed a total of $20,000, and community service for not more than 500 hours.[243]

For animal fighting, the most egregious acts are felonies punishable by imprisonment for not more than four years, a fine of not less than $5,000 or more than $50,000, and not less than 500 or more than 1,000 hours of

235. *Id.* at *3.
236. MICH. COMP. LAWS § 750.50 (2004 & Supp. 2010).
237. MICH. COMP. LAWS § 750.50(4)(a) (2004 & Supp. 2010).
238. *Id.* § 750.50(4)(b).
239. *Id.* § 750.50(4)(c).
240. *Id.* § 750.50(4)(d).
241. *Id.* § 750.50(6).
242. *See id.* § 750.50.
243. *Id.* § 750.50(4).

community service.[244] A person who commits a less egregious violation of the animal fighting statute is also guilty of a felony punishable by imprisonment for not more than four years, but receives a lesser fine and fewer hours of community service.[245] A person can be charged with, convicted of, or punished for any other violation of law committed while violating the animal fighting section.[246]

Michigan's law is laudable. It does not differentiate between species of animals and protects them by imposing serious penalties on their abusers. In cases of neglect, Michigan enhances sentences when there are prior convictions and multiple injured animals.[247] By recommending penalties, Michigan removes considerable prosecutorial and judicial discretion in charging and sentencing.[248]

Michigan did not abandon centuries of jurisprudence establishing that defendants can be charged for multiple violations of the same statute only in cases involving human victims, but instead creatively enhanced the sentences of certain types of animal abusers. Michigan's law applies both property and human victim principles. The human victim principles Michigan adopts are: (1) like human victims, there is no distinction among animal victims; (2) perpetrators receive greater punishment if more animals are injured; and (3) sentences are enhanced if there are prior convictions.[249] The property principles contained in Michigan's law are: (1) perpetrators who commit intentional acts do not receive greater punishment if multiple animals are involved; and (2) perpetrators who commit intentional acts and have prior convictions do not receive an enhanced punishment.[250]

Michigan's law regulating intentional acts is inadequate because it imposes no additional punishment for prior animal abuse convictions or multiple injured animals. It also fails to authorize additional convictions or punishments for other violations of law arising out of the same transaction and does not permit consecutive sentences. Additionally, Michigan's statute regulating negligent abuse of animals fails to protect injured animals in quantities over ten.[251]

---

244. *Id.*
245. *Id.* § 750.49.
246. *Id.*
247. *Id.* § 750.50.
248. *Id.*
249. *Id.*
250. *Id.* § 750.50(4).
251. *Id.*

V. RECOMMENDATIONS

A. The Human Victim, Property, Living Property, Personhood, or Equal Protection Model?

Whether or not multiple violations of the same criminal statute constitute separate offenses should depend upon the plain language of a statute. If a legislature has expressly authorized the charging of multiple counts for multiple violations of a single act and has drafted its statute in a way that clearly reflects that intention, then it is a valid statute that does not constitute a violation of due process or the double jeopardy clause.[252] If such legislative intent is not apparent, then the rule of lenity requires that only a single count be charged.[253]

Yet, as this article has demonstrated, prosecutors in states that have failed to address the issue of charging when multiple animals are injured at the same time are charging defendants with multiple counts. The forty-one states without legislation addressing multiple counts should enact statutes specifically covering the issue.[254] As we have seen, Wyoming,[255] Ohio,[256]

---

252. Robinson, *supra* note 159, at § 68.

253. The United States Supreme Court has ruled that, if the scope of permissible punishment in a criminal statute is unclear or ambiguous from the statutory language, and there is insufficient information about legislative intent to resolve that ambiguity, the rule of lenity applies, prohibiting the court from imposing the harsher punishment. *See* Bell v. United States, 349 U.S. 81, 83 (1955) (The Federal Mann Act did not fix the exact punishment for transporting multiple people across state lines for immoral purposes; accordingly, defendant could only be convicted on one count of violating the statute, even though he transported two people.). *See generally* H. Mitchell Caldwell & Jennifer Allison, *Counting Victims and Multiplying Counts: Business Robbery, Faux Victims, and Draconian Punishment*, 46 IDAHO L. REV. 647 (2010) (discussing the number of counts that may be applied to a single incident).

254. *See supra* note 150; *see also supra* Part III.A.2. California, Pennsylvania, Georgia, Michigan, and New Jersey should enact statutes broader than the ones currently in place. The legislation in those states is inadequate to fully protect animals.

255. Wyoming Statutes Annotated section 6-3-203 (1977) provides, in relevant part:

> (a) A person commits cruelty to animals if he knowingly and with intent to cause death, injury or undue suffering:
>
> (i) Overrides an animal or drives an animal when overloaded; or
>
> (ii) Unnecessarily or cruelly beats, tortures, torments, injures, mutilates or attempts to kill an animal; or
>
> (iii) Carries an animal in a manner that poses undue risk of injury or death.
>
> (b) A person commits cruelty to animals if he has the charge and custody of any animal and unnecessarily fails to provide it with the proper food, drink or protection from the weather, or cruelly abandons the animal, or in the case of immediate, obvious, serious illness or injury, fails to provide the animal with appropriate care.

and California[257] have statutes that are mostly silent on the issues of charging and sentencing when multiple animals are injured at the same time. Courts in these states have already considered appeals in which

---

WYO. STAT. ANN. § 6-3-2003 (1977).

256. Ohio Revised Code Annotated section 959.13 (West 1976) provides, in relevant part:

> (A) No person shall:
>
> (1) Torture an animal, deprive one of necessary sustenance, unnecessarily or cruelly beat, needlessly mutilate or kill, or impound or confine an animal without supplying it during such confinement with a sufficient quantity of good wholesome food and water;
>
> (2) Impound or confine an animal without affording it, during such confinement, access to shelter from wind, rain, snow, or excessive direct sunlight if it can reasonably be expected that the animal would otherwise become sick or in some other way suffer. Division (A)(2) of this section does not apply to animals impounded or confined prior to slaughter. For the purpose of this section, shelter means a man-made enclosure, windbreak, sunshade, or natural windbreak or sunshade that is developed from the earth's contour, tree development, or vegetation.

OHIO REV. CODE ANN. § 959.13 (West 1976).

257. California Penal Code section 597 provides, in relevant part:

> (a) Except as provided in subdivision (c) of this section or Section 599c, every person who maliciously and intentionally maims, mutilates, tortures, or wounds a living animal, or maliciously and intentionally kills an animal, is guilty of an offense punishable by imprisonment in the state prison, or by a fine of not more than twenty thousand dollars ($20,000), or by both the fine and imprisonment, or, alternatively, by imprisonment in a county jail for not more than one year, or by a fine of not more than twenty thousand dollars ($20,000), or by both the fine and imprisonment.
>
> (b) Except as otherwise provided in subdivision (a) or (c), every person who overdrives, overloads, drives when overloaded, overworks, tortures, torments, deprives of necessary sustenance, drink, or shelter, cruelly beats, mutilates, or cruelly kills any animal, or causes or procures any animal to be so overdriven, overloaded, driven when overloaded, overworked, tortured, tormented, deprived of necessary sustenance, drink, shelter, or to be cruelly beaten, mutilated, or cruelly killed; and whoever, having the charge or custody of any animal, either as owner or otherwise, subjects any animal to needless suffering, or inflicts unnecessary cruelty upon the animal, or in any manner abuses any animal, or fails to provide the animal with proper food, drink, or shelter or protection from the weather, or who drives, rides, or otherwise uses the animal when unfit for labor, is, for every such offense, guilty of a crime punishable as a misdemeanor or as a felony or alternatively punishable as a misdemeanor or a felony and by a fine of not more than twenty thousand dollars ($20,000).

CAL. PENAL CODE § 597 (West 2011).

prosecutors filed charges against a defendant for more than one animal injured in the same incident.[258] The prosecutors in these three states filed charges in a myriad of ways. In Wyoming, forty horses and forty-five cows were mistreated; yet, the defendant was charged with nine counts of animal cruelty—six counts referred to horses and three to cows.[259] In Ohio, twenty-eight cows were mistreated and twenty-eight counts were filed.[260] In California, the charging decisions displayed a lack of uniformity. Twelve injured geese, one injured chicken, two injured dogs, multiple dead and dying calves and chickens, dead cows, many dogs without food or water, four or five dead rabbits, a dead peacock, one or two dead pigs, and one injured puppy were found on the defendant's property.[261] Somehow, nine counts of animal cruelty were filed.[262] One count involved an injured dog, and another a severely wounded puppy. The injured rabbits, ducks, chickens, and geese were the basis of the remaining counts.[263] The prosecutor lumped some animals together by species and charged the defendant with separate offenses for some animals alone.[264]

These three courts reached different results based on identical statutory schemes.[265] None of these states authorized its prosecutors to file one count per animal,[266] yet prosecutors in all of these cases did.[267] The subsequent discrepancy in charging decisions reveals the importance of clear statutes that do not leave important charging decisions to prosecutors' discretion.

But how should animal abuse legislation read? What approach is best? If the objective of animal cruelty statutes is the protection of animals, then the criminal act should be treated more like a crime against the person than a crime against property. Precedent certainly exists. Many jurisdictions prosecute robbery—which is a hybrid crime against the person (the assault) as well as property (the theft)—using the person assaulted or robbed and not the property taken as the "unit of prosecution."[268] In these jurisdictions, multiple counts can be filed when there are multiple victims involved in a robbery. In contrast, in other jurisdictions, only one count can be filed in this situation, treating robbery as a property crime.[269] Though the law has

---

258. *See supra* Part IV.B.

259. *See* Amrein v. State, 836 P.2d 862, 863 (Wyo. 1992).

260. *See* State v. Lapping, 599 N.E.2d 416, 417–18 (Ohio Ct. App. 1991).

261. People v. Sanchez, 114 Cal. Rptr. 2d 437, 440–41 (Cal. Ct. App. 2001).

262. *Id.* at 443.

263. *Id.* at 446–47.

264. *Id.* at 446.

265. *See supra* Part IV.B.

266. *See supra* Part IV.B.

267. *See generally Sanchez*, 114 Cal. Rptr. 2d 437 (charging one count per animal); State v. Lapping, 599 N.E.2d 416 (Ohio Ct. App. 1991) (charging one count per animal); Amrein v. State, 836 P.2d. 862 (Wyo. 1992) (charging one count per animal).

268. People v. Wakeford, 341 N.W.2d 68, 73 (Mich. 1983).

269. *See* Caldwell, *supra* note 252, at 125.

historically viewed animals as property, they are also living creatures that should be protected against some, if not all, physical mistreatment.[270]

Legislatures have responded to this change in society's humane demands by enacting laws that protect animals, and some courts are even viewing animals like human victims.[271] Accordingly, it is appropriate to conclude that the proper unit of prosecution for crimes of violence against animals should be the animals assaulted or killed and not the underlying criminal act that violates another person's property interest in his or her animals.[272]

Animal rights scholars are proposing new theories to resolve the moral dilemma presented by the fact that animals' rights are subordinate to human rights in the crafting of animal rights protections. Animal rights scholar Gary Francione argues that nonhuman animals should be treated as persons.[273] There is precedent for this fiction:

> It is well established that being human is not essential to obtaining 'personhood.' For example, corporations have long been considered 'persons' under the Fourteenth Amendment . . . . And ships historically have been provided legal status that might be the envy of modern animal advocates . . . . These are legal fictions that do not recognize any inherent qualities in the corporation or ship, but simply acknowledge the need to bring issues involving them before the courts.[274]

A move toward personhood would entail a dramatic departure from the current legal treatment of animals as property.[275] As persons, animals would possess legal rights to avoid suffering and to live and enjoy a protected existence.[276]

Another proposal, by legal philosopher Ani Satz, advocates the creation of the "Equal Protection of Animals (EPA) paradigm."[277] Satz argues that "interest convergence," the phenomenon of a privileged group providing legal protections to a disadvantaged group when providing such protections supports the interests of the privileged group, applies to animals because they receive legal protections only when their interests align with human interests.[278] Satz states:

---

270. *See* Chiesa, *supra* note 2, at 24–39.

271. *See supra* Part II.

272. Commonwealth v. Melton, 763 N.E.2d 1092, 1098 (Mass. 2002).

273. FRANCIONE, *supra* note 26, at 14.

274. SONIA S. WAISMAN ET AL., ANIMAL LAW 73 (2d ed. 2002).

275. *See* FRANCIONE, *supra* note 26, at 102; *see also* Taimie L. Bryant, *Sacrificing the Sacrifice of Animals: Legal Personhood for Animals, the Status of Animals as Property, and the Presumed Primacy of Humans*, 39 RUTGERS L.J. 247, 247–48 (2008).

276. *Id.*

277. Satz, *supra* note 84, at 73.

278. *Id.* at 67–69. For example, Satz states that the only distinction between the legal treatment of the pig and the dog is human emotional attachment. *Id.* at 98.

> Daily consumer choices deny that the capacity of animals to suffer is morally relevant. The reason for this is obvious: If animal capacities are morally relevant, current use of domestic animals for food, experimentation, exhibition, entertainment, and some forms of service must end, or people must acknowledge engaging in daily, immoral practices.[279]

Under the EPA paradigm, Satz propounds:

> [D]omestic animals must have the ability to intake necessary food and hydration, have necessary shelter and exercise and be able to engage in natural behaviors of movement, maintain bodily integrity (including avoiding pain inflicted on the body), and experience companionship. EPA requires that these basic capabilities of animals—human and nonhuman—are maximized within a given population . . . . This entails a shift from the presumption that animals may be used for human purpose with some restrictions to a presumption against animal use absent justification. Animal use for human purpose is justified only if it also maximizes the enumerated basic capabilities for animals.[280]

Under this theory, "[f]actory farms would be abolished, though one could imagine some small-scale farming operations," and "it is doubtful that most animal experimentation could continue."[281]

Perhaps the paradigms advocated by Francione and Satz will be implemented over time if our society becomes more interested in the welfare of animals. At the present time, however, even Satz recognizes that when humans seek to improve the conditions of animals, they are prompted to do so because of human need rather than a desire to help animals.[282] Therefore, it is unlikely that these paradigms will be implemented anytime soon.

Lastly, animal rights scholar David Favre advocates treating animals as "living property." He believes that animals are neither human nor inanimate objects:

> Presently, the law has only two clearly separated categories: property or juristic person. But, by using existing concepts of property law, it is possible to construct a new paradigm that gives animals the status of juristic persons without entirely severing the concept of property ownership. It is a blending of the two previously separated categories.[283]

---

279. *Id.* at 101.

280. *Id.* at 73.

281. *Id.* at 116.

282. *Id.* at 84.

283. David Favre, *Equitable Self-Ownership for Animals*, 50 DUKE L.J. 473, 502 (2000); *see generally* David Favre, *Living Property: A New Status for Animals within the Legal System*, 93 MARQ. L. REV. 1021 (2010) (proposing that non-human animals should be

This expanded property status may not resolve moral or legal inconsistencies resulting from the unequal treatment of animal suffering, because recognizing animals as living property does not necessarily translate into equal treatment. It is, however, a good compromise. Many people will balk at treating animals in the same way, legally, as human victims. However, treating animals like property is somewhat ludicrous given how modern society views animals as living creatures that need protection and can feel pain. Establishing this third "living property" status could provide an attractive alternative for people on either side of the issue because it satisfies our social need to protect animals but does not elevate them to the same status as humans.

The legal personhood, expanded or "living" property, and EPA paradigms rely on a change in legal status to generate greater protection for animals. Effecting this change is a complicated process requiring the cooperation and continued efforts of scholars and practitioners of various legal backgrounds. This article, however, centers on animal cruelty, not individual rights to animals.

The statute proposed in Appendix A would work to affect an improvement to the status of animals without relying on a change in legal status. Each legally recognizable "victim" would constitute the basis of one count. The statute proposed is a more pragmatic approach than is currently available within our existing system of criminal jurisprudence. Although it is incremental in nature, it can at least grant additional animals immediate protection. Similar language has been implemented in Alaska, Wyoming, Louisiana, Montana, and Arkansas without any indication of problems.[284] Although the proposed statute does nothing to protect animals currently exempted from animal abuse legislation, at least it would increase the punishment of those who hurt more than one animal at the same time, immediately improving the rights of more animals than at present. Hopefully, enhanced sentences will deter the commission of abuse. Increased accountability may help to change societal perceptions about the value of animals.[285]

Regardless, it is reasonable to punish offenders for the injury and death of each harmed animal. However, there are valid issues and concerns. Animals, after all, have been historically perceived as property, are purchased and sold, are dominated by humans, and cannot always be identified from one another. Further, since farm animals are eventually slaughtered and eaten, and wild animals can be hunted and killed, is there a policy that makes sense?

---

granted legal rights and explaining how property law principles may be adjusted to accommodate the recognition of animal ownership as a new category of property).

284. *See supra* Table I. This issue has not been litigated in any of these states.

285. *See supra* Part III.B.

B. Hypotheticals

The following hypotheticals reveal but one legally sound solution presently available within our existing system of criminal jurisprudence. States should follow the lead of Alaska, Wyoming, Louisiana, Montana, and Arkansas and apply human victim principles to prosecutions involving multiple animals.

1. If three dogs are found to be starving and only one food bowl is located, should a prosecutor be able to charge the defendant with one or three counts of animal abuse? Under human victim charging and sentencing rules, the defendant could be charged with three counts of abuse because there are three "victims"; under property rules, however, only one count would be permitted, as the crimes are not separated in time or place.

If the same three starving dogs are recovered from three separate, locked cages each containing a food bowl, should the defendant be charged with one or three offenses? Under either a human victim or property-based analysis, the prosecutor should charge the defendant with three counts of abuse because each offense is separate with its own possible defenses. It does not make sense to differentiate these factual situations. Three animals are abused under either scenario. Only if we apply human victim principles do we get the same result.

2. If three animals are intentionally poisoned and die, and only one bowl containing poison is located, should the defendant be charged with one or three offenses? Applying a property analysis, because the abuse is pursuant to one criminal objective, multiple counts are not permissible. If the animals ate out of the same bowl of poison at the same time, only one count could be brought against the defendant.

If the same three animals are separately poisoned and three bowls containing poison are found, under either a human victim or a property-based analysis, the defendant could be charged with three separate counts.[286] Only under a human victim analysis would the result be the same whether there were one or three bowls of poison; the three dead animals would form the basis of three separate counts. As with the example of negligent failure to feed, logically the result should be the same in each factual situation.

3. A defendant raising multiple animals for fighting could either have one objective (the intention to raise dogs for fighting), or he could have multiple criminal objectives (the intention to raise dogs for multiple and separate fights). If a property analysis is applied, the result would vary. If a defendant has the single intent and objective to raise dogs for fighting, the defendant could be charged with one count; if a defendant has a separate intent and objective as to each animal, the defendant could be charged with multiple counts.

---

286. In one Idaho case, acts of administering or exposing poison were in fact charged together. State v. Farnsworth, 10 P.2d 295, 297 (Idaho 1932).

If animals raised for fighting are treated like human victims, each injured animal could result in a separate charge in either scenario.[287] This is a better approach, as it seems wrong to punish a defendant in possession of multiple, injured dogs clearly raised for fighting for only one offense.

The answers to these three scenarios should be the same. The analysis reveals that, in the criminal law context, a property-based approach produces inequitable results when multiple animals are abused.[288] What would be the result under a living personhood, equal protection, or living property paradigm? Under the personhood or equal protection paradigm, the number of counts that could be filed would be the same in each of the three scenarios. Each harmed animal would form the basis of one count regardless of the number of food bowls, cages, poison, or fights. Furthermore, no animal would be excluded. All animals, regardless of status or activity, would be protected at all times. In essence, the personhood or equal protection model would concentrate on the harm to the animal rather than the effect of the harm on people. Each animal would form the basis for a unit of prosecution.

Under a living property paradigm, the result is unclear. The living property paradigm is based on the legal obligation that the owner of an animal has to the animal, "thus creating a relationship closer to the nature of a guardianship."[289] Therefore, if an animal is harmed by its owner in any of the above-described hypothetical situations, each animal could form the basis of a charge. However, if an animal is harmed by someone other than the owner, traditional property rules might apply.

At this point, however, the personhood, equal protection and living property paradigms are simply subjects of scholarly discussion. The only solution that is presently available and that appears to be logical and legally sound is to modify state statutes to hold defendants responsible for each animal that they injure.

C. What About the Michigan Model?

What about Michigan's hybrid sentencing scheme applying both property and human victim principles? At first glance, it seems like a good idea. It takes into account injuries to multiple animals and attempts to protect animals without modifying their historical perception as property.[290]

---

287. *See* OR. REV. STAT. § 161.067(2) (2009).

288. Compare this to tort law where matters such as capacity to sue and standing are significant issues. *See generally* Richard L. Cupp, Jr., *A Dubious Grail: Seeking Tort Law Expansion and Limited Personhood as Stepping Stones Toward Abolishing Animals' Property Status*, 60 SMU L. REV. 3, 24 (2007) (advocating standing for animals as plaintiffs). But in the criminal law context, the parties are the state and the defendant, and issues such as an animal's capacity to sue and standing are not present.

289. ANIMAL LAW WELFARE, *supra* note 11, at 431.

290. MICH. COMP. LAWS § 750.50(b) (2004).

However, the law falls short because of its failure to conform to our current system of charging. The Michigan Legislature has, in effect, created a third-tier for animals. Of course, there are other hybrid statutes and offenses—robbery, for example, is in some ways a property crime and in other ways a human victim offense.[291] But courts and legislators considering the issue of multiple robbery victims have not created an entirely separate tier for robbery. Instead, some states permit a defendant to be sentenced for each robbery victim, while others treat robbery as a property crime, permitting the defendant to be sentenced for only one offense.[292]

Michigan's law also fails to address the question of what should happen when more than ten animals are negligently abused.[293] Furthermore, Michigan fails to address the injuries to individual negligently treated animals within each of the sub-groups—each capable of experiencing pain.[294] Instead, Michigan groups animals together in much the same way as items of property. For example, if a perpetrator injures nine animals, he receives the same punishment as if he injures four animals. There is no precedent for grouping of persons or property in the filing of complaints. Although the Michigan model is better than the property model, it is legally flawed and inadequate. The only way to resolve the issue of multiple counts of animal abuse in a consistent and legally sound manner is to consider each injured animal a separate count.

Applying human victim principles to animal cruelty law would mean that each abused animal would constitute a separate offense and that all animals would be treated in the same manner. Moreover, like the rest of our system of criminal law, statutes could provide exceptions excusing harm done to animals if it is legally justified.[295] Even if human victim principles are adopted, our laws can continue to grant exceptions when animals are (1) raised for slaughter, as long as they are not starved or tortured in the process;[296] (2) shot by hunters, in accordance with state and federal regulations;[297] (3) killed for scientific research, as long as the laboratories are in compliance with state and federal guidelines; and (4) harmed by many other activities currently permitted by state legislatures throughout

291. *See* OR. REV. STAT. § 161.067 (2009).

292. *See* Caldwell, *supra* note 252, at 125; *see also* OR. REV. STAT. § 161.067(2) ("When the same conduct or criminal episode, though violating only one statutory provision involves two or more victims, there are as many separately punishable offenses as there are victims."). *See generally* State v. Williams, 209 P.3d 842 (Or. Ct. App. 2009) (interpreting Oregon Revised Statutes section 161.067).

293. MICH. COMP. LAWS § 750.50 (2010).

294. *Id.*

295. Ignorance or mistake, entrapment, duress, necessity, public duty, self-defense, defenses of another, defense of property, law enforcement, consent by victim are some of the justifications and excuses to harm for harm done to human victims. LAFAVE, *supra* note 68, at 575–76.

296. 7 U.S.C.A. § 1902 (2010).

297. 16 U.S.C.A. § 460o-5 (2010).

our country.[298] Adopting the human victims approach would not upset any of our principles governing permissible human behavior towards animals. In fact, the human victims approach is already being applied in Alaska, Wyoming, Louisiana, Montana, and Arkansas.[299] Following the lead of these five progressive states will add much-needed consistency and simplicity to the law on animal abuse.

D. A Proposed Statute

An animal abuse statute should include separate provisions for intentional acts, negligent abuse, and animal fighting. It should authorize felony charges for intentional abuse and animal fighting. It should provide for increased sentences for more severe abuses and should recommend maximum sentences for each violation. It should remove animals from the care of repeat offenders. It should authorize separate counts for each injured animal, and consecutive sentences should be permissible if there are other violations of law arising out of the same transaction.

A proposed statute is set forth in Appendix A. It is based in part on Michigan's law, in that it (1) adopts Michigan's system of separating negligent, intentional, and dog fighting acts; (2) includes recommended sentences; (3) adopts Michigan's "arising out of the same transaction" language; (4) includes Michigan's provision removing animals from the care of perpetrators under certain circumstances; and (5) contains a consecutive sentencing provision.[300] It also follows Michigan's lead in dividing animal fighting into two categories with different sentencing recommendations for each.[301] When multiple animals are injured at the same time, it follows the lead of Alaska, Wyoming, Louisiana, Montana, and Arkansas by including a provision providing that each injured animal shall constitute a separate count for negligent, intentional, and dog fighting acts.[302]

---

298. The Animal Welfare Act was signed into law in 1966. While Congress' original intent was to regulate the care and use of animals in the laboratory, the Act has become the only federal law in the United States that regulates the treatment of animals in research, exhibition, transport, and by dealers. The Animal Welfare Act sets forth the minimum acceptable standard. The Act has been amended four times and can be found in *U.S.C.A.* §§ 2131–56 (2006). The Act is enforced by the U.S. Department of Agriculture, Animal and Plant Health Inspection Service (APHIS), and Animal Care (AC). For more information, see the Animal Welfare Information Center's website at http://awic.nal.usda.gov.

299. *See supra* note 126 and accompanying text.

300. *See* MICH. COMP. LAWS § 750.50(b).

301. *See id.*

302. *See supra* note 126 and accompanying text.

V. CONCLUSIONS

Mahatma Gandhi once said that "[t]he greatness of a nation and its moral progress can be judged by the way its animals are treated."[303] This country has struggled mightily to find a way to protect its animals while also maintaining its right to eat, test, and kill them. Our country has slowly transitioned from a society with little concern for the pain endured by animals to a culture that occasionally thinks about its chickens laying eggs, the lab animals testing its drugs, and its methods of butchering animals. The fact that our society uses animals for sport, food, and science has perhaps made it more difficult to extend greater protection to animals. For how can we treat them like people when we are permitted to kill them?

We can continue our present uses of animals and still extend greater protection to animals under the law. The framework already exists. Laws pertaining to animals already set forth offenses and then list exceptions. Already, we may legally kill animals for certain "justified" reasons, just as it is legally acceptable to kill humans under certain circumstances such as necessity or duress.

States should be encouraged to modify their statutes to provide specifically that each injured animal should constitute a separate count and that all animals are entitled to protection under our laws unless a legally acceptable excuse is proffered.

Until we decide to treat each injured animal as a separate being, prosecutors, judges, and legislatures will continue to bend and distort our laws to reach the results they desire. As society seems to be increasingly offended by the way animals are treated by some of its members, this result-oriented approach may produce favorable outcomes. However, the statute proposed by this paper seeks to add more consistency to the law. Beyond an individual result at trial, this paper advocates for a paradigm shift in which each animal is perceived as a victim itself.

For example, a drunk driver recently crashed into a van, killing five people, and was charged with five counts of criminally negligent homicide.[304] What if the drunk driver instead crashed into an animal control vehicle and injured ten dogs? The result should not be any different—ten counts should be charged. Clearly, as stated by Justice Stevens, humans are far more worthy of protection in our society,[305] but the difference in result is not logical. If a man abuses his children and pets at the same time, why should it be acceptable for him to be charged separately for his children's injuries but not for his pets' injuries? Until we can see the legal

303. Addie Patricia Asay, Note, *Greyhounds: Racing to Their Deaths*, 32 STETSON L. REV. 433, 464 (2003) (quoting Mahatma Gandhi, The Moral Basis of Vegetarianism, Address to the London Vegetarian Society (Nov. 20, 1931)).

304. *Crash that Killed 5 Amish Took Hours to Untangle*, CBS NEWS (July 7, 2011), http://www.cbsnews.com/stories/2011/07/20/earlyshow/main20081014.shtml.

305. United States v. Stevens, 130 S. Ct. 1577, 1600 (2010).

inconsistency in applying a different result to these questions, we will continue to experience illogical charging outcomes that sometimes join all animals together and sometimes do not. For "the question is not, Can they reason? nor, Can they talk? but, Can they suffer?"[306] Yes, each of them can suffer, and each one of them should be entitled to legal protection. We, as a society, must decide that their abuse is unacceptable.

306. JEREMY BENTHAM, AN INTRODUCTION TO THE PRINCIPLES OF MORALS AND LEGISLATION ch. 17 n. 122 (2d ed. 1823), *available at* http://www.econlib.org/library/Bentham/bnthPML.html.

APPENDIX A

PROPOSED STATUTE

1. Duty of Care:
    a. An owner, possessor, or person having the charge or custody of an animal shall not do any of the following:
        1) Fail to provide an animal with adequate care.
        2) Cruelly drive, work, or beat an animal.
        3) Abandon an animal.
        4) Negligently allow any animal to suffer unnecessary neglect, torture or pain.
    b. A person who violates subsection (a) is guilty of a misdemeanor punishable by one or more of the following:
        1) Imprisonment for not more than one year.
        2) A fine of not more than $2,000.00.
        3) Community service for not more than 300 hours.
    c. Upon a third conviction, a person shall lose his or her right to own and care for any animal.
    d. Each animal injured pursuant to this section shall constitute a separate count.
    e. This section does not prohibit a person from being charged with, convicted of, or punished for any other violation of law arising out of the same transaction as the violation of this section.
    f. The court may order a term of imprisonment imposed for a violation of this section to be served consecutively to a term of imprisonment imposed for any other crime including any other violation of law arising out of the same transaction as the violation of this section.
2. Intentional Abuse:
    a. A person shall not do any of the following without just cause:
        1) Knowingly kill, torture, mutilate, maim, or disfigure an animal.
        2) Commit a reckless act knowing or having reason to know that the act will cause an animal to be killed, tortured, mutilated, maimed, or disfigured.
        3) Knowingly administer poison to an animal, or knowingly expose an animal to any poisonous substance, with the intent that the substance be taken or swallowed by the animal.

b. A person who violates subsection (a) is guilty of a felony punishable by one or more of the following:
    1) Imprisonment for not more than four years.
    2) A fine of not more than $10,000.
    3) Community service for not more than 500 hours.
c. Upon a conviction of this section, a person loses his right to own or care for any animal.
d. Each animal that is abused pursuant to this section shall constitute a separate offense.
e. This section does not prohibit a person from being charged with, convicted of, or punished for any other violation of law arising out of the same transaction as the violation of this section.
f. The court may order a term of imprisonment imposed for a violation of this section to be served consecutively to a term of imprisonment imposed for any other crime including any other violation of law arising out of the same transaction as the violation of this section.

3. Animal Fighting:
    a. A person shall not knowingly do any of the following:
        1) Own, possess, use, buy, sell, offer to buy or sell, import, or export an animal for fighting; be a party to or cause the fighting, baiting, or shooting of an animal; rent or otherwise obtain the use of a building, shed, room, yard, ground, or premises for fighting, baiting, or shooting an animal; or permit the use of a building, shed, room, yard, ground, or premises belonging to him or her or under his or her control for any of the purposes described in this section.
        2) Organize, promote, or collect money for the fighting, baiting, or shooting of an animal as described in subdivision (1); be present at a building, shed, room, yard, ground, or premises where preparations are being made for an exhibition; or breed, buy, sell, offer to buy or sell, exchange, import, or export an animal the person knows has been trained or used for fighting.
    b. A person who violates subsection (a)(1) is guilty of a felony punishable by one or more of the following:
        1) Imprisonment for not more than four years.
        2) A fine of not less than $5,000.00 or more than $50,000.00.

3) Not less than 500 or more than 1,000 hours of community service.

c. A person who violates subsection (a)(2) is guilty of a felony punishable by one or more of the following:
   1) Imprisonment for not more than two years. A fine of not less than $1,000.00 or more than $5,000.00.
   2) Not less than 250 or more than 500 hours of community service.
d. Upon a conviction of subsection (a), a person loses his right to own or care for any animal.
e. Each act described in subsection (a) shall constitute a separate offense.
f. This section does not prohibit a person from being charged with, convicted of, or punished for any other violation of law that is committed by that person while violating this section.

TENNESSEE LAW REVIEW

ANNUAL INDEX
VOLUME 78
2010-2011

THE UNIVERSITY OF TENNESSEE
COLLEGE OF LAW

INDEX TO VOLUME 78

AUTHOR INDEX 1173
SUBJECT INDEX 1177
TABLE OF CASES 1179

Tennessee Law Review Association, Inc.
1505 West Cumberland Avenue
Knoxville, Tennessee 37996-1810

AUTHOR INDEX

BARNETT, RANDY E.: *The Case for the Repeal Amendment* 813

BARTOW, ANN: *An Equal Rights Amendment to Make Women Human* 839

BLAZE, DOUG: *Semper Fidelis* 637

CALVERT, DAN: *Tort Law—Comparative Fault—Original Tortfeasor Rule in Tennessee* 259

CARTER, JEFF: *Strictly Business: A Historical Narrative and Commentary on Rock and Roll Business Practices.* 213

DENNING, BRANDON P.: *Three Cheers for Bob Lloyd* 639

DENNING, BRANNON P. & SMITH, BROOKS R.: *The Truth in Legislation Amendment: An Idea Whose Time Has Come* 831

EDITORS OF THE TENNESSEE LAW REVIEW: *Article V Constitutional Conventions: A Primer* 663

ELAM, CAITLYN LUEDTKE: *Constitutional Law—Fifth Amendment—Honing the Protections Offered by Miranda Through Supreme Court Bright-Line Rulemaking* 569

ELLIS, SARA R.: *Copyrighting Couture: An Examination of Fashion Design Protection and Why the DPPA and IDPPPA are a Step Towards the Solution to Counterfeit Chic* 163

EPSTEIN, RICHARD A.: *Why We Need Term Limits for Congress: Four in the Senate, Ten in the House* 849

FOLEY, ELIZABETH PRICE: *Sovereignty, Rebalanced: The Tea Party & Constitutional Amendments* 751

GALLIGAN, JR., THOMAS C.: *Pilot Lloyd* 643

GIANNINI, MARY MARGARET: *Redeeming an Empty Promise: Procedural Justice, the Crime Victims' Rights Act, and the Victim's Right to be Reasonably Protected from the Accused* 47

HADFIELD, CHELSEY J.: *Civil Procedure—Intervention—Nonsovereign Entities in Equitable Apportionment Actions Invoking Original Jurisdiction* 613

HEMINWAY, JOAN MACLEOD: *What About Bob?* 649

HEMINWAY, JOAN MACLEOD & HOFFMAN, SHELDEN RYAN: *Proceed at Your Peril: Crowdfunding and the Securities Act of 1933* 879

JONES, JEFFREY DOUGLAS: Property Rights, Property Wrongs, and Chattel Dispossession Under Self-Storage Leases 1015

LAT, DAVID & SHEMTOB, ZACHARY: *The Execution Should be Televised: An Amendment Making Executions Public* 859

LEVINSON, SANDFORD: *Afterword: Full of Sound and Fury but Signifying Relatively Little?* 867

LYNCH, TIMOTHY: *Amending Article V to Make the Constitutional Amendment Process Itself Less Onerous* 823

MALIK, YUSUF Z.: *The Religious Land Use and Institutionalized Persons Act: A Perspective on the Unreasonable Limitations Provision* 531

MANNING, MORGAN LEIGH: *Less Than Picture Perfect: The Legal Relationship Between Photographers' Rights and Law Enforcement* 105

MCFARLAND, IAN G.: *In the Wake of* Therasense & Nisus Corp., *How Can Patent Attorneys Defend Themselves Against Allegations of Inequitable Conduct* 487

MCLEOD, SCOTT M.: *Constitutional Law—Sixth Amendment Right to Counsel—Forfeiture of the Right to Counsel in Tennessee* 589

MULROY, STEVEN J.: *Sunlight's Glare: How Overbroad Open Government Laws Chill Free Speech and Hamper Effective Democracy* 309

NATELSON, ROBERT G.: *Proposing Constitutional Amendments by Convention: Rules Governing the Process* 693

OLIVER, WESLEY MACNEIL: *The Modern History of Probable Cause* 377

PENROSE, MARY MARGARET: *Conventional Wisdom: Acknowledging Uncertainty in the Unknown* 789

PERDUE, ABIGAIL LAUREN: *Justifying GINA* 1051

REYNOLDS, GLENN HARLAN: *Foreword: Divine Operating System* 651

RICHARDSON, SALLY BROWN: *Nonuse and Easements: Creating a Pliability Regime of Private Eminent Domain* 1

ROBINETTE, CHRISTOPHER J.: *Why Civil Recourse Theory is Incomplete* 431

SERAFINO, LAURIE: *No Walk in the Dog Park: Drafting Animal Cruelty Statutes to Resolve Double Jeopardy Concerns and Eliminate Unfettered Prosecutorial Discretion* 1119

SOLED, JAY A. & GANS, MITCHELL: *Sales to Grantor Trusts: What the IRS and Congress Can Do to Curb Aggressive Transfer Tax Techniques* 973

STERN, MICHAEL: *Reopening the Constitutional Road to Reform: Toward a Safeguarded Article V. Convention* 765

STERN, MICHAEL: *A Brief Reply to Professor Penrose* 807

THOMPSON, DOROTHEA K.: *Bankrutpcy Law—Dischargeability of Student Loan Debt—Application of Federal Rule of Civil Procedure 60(B)(4) to Discharge-By-Declaration Provisions* 287

SUBJECT INDEX

BANKRUPTCY LAW

Bankruptcy Law—Dischargeability of Student Loan Debt—Application of Federal Rule of Civil Procedure 60(B)(4) to Discharge-By-Declaration Provisions 287

CIVIL PROCEDURE

Civil Procedure-Intervention-Nonsovereign Entities in Equitable Apportionment Actions Invoking Original Jurisdiction 613

CONSTITUTIONAL LAW

Afterword: Full of Sound and Fury but Signifying Relatively Little? 867

Amending Article V to Make the Constitutional Amendment Process Itself Less Onerous 823

Article V Constitutional Conventions: A Primer 663

A Brief Reply to Professor Penrose ..807

The Case for the Repeal Amendment 813

Constitutional Law-Fifth Amendment-Honing the Protections Offered by Miranda Through Supreme Court Bright-Line Rulemaking 569

Constitutional Law-Sixth Amendment Right to Counsel-Forfeiture of the Right to Counsel in Tennessee 589

Conventional Wisdom: Acknowledging Uncertainty in the Unknown 789

An Equal Rights Amendment to Make Women Human 839

The Execution Should be Televised: An Amendment Making Executions Public 859

Foreword: Divine Operating System 651

Less Than Picture Perfect: The Legal Relationship Between Photographers' Rights and Law Enforcement 105

Proposing Constitutional Amendments by Convention: Rules Governing the Process 693

The Religious Land Use and Institutionalized Persons Act: A Perspective on the Unreasonable Limitations Provision 531

Reopening the Constitutional Road to Reform: Toward a Safeguarded Article V Convention 765

Sovereignty, Rebalanced: The Tea Party & Constitutional Amendments 751

Sunlight's Glare: How Overbroad Open Government Laws Chill Free Speech and Hamper Effective Democracy 309

The Truth in Legislation Amendment: An Idea Whose Time Has Come 831

Why We Need Term Limits for Congress: Four in the Senate, Ten in the House 849

COPYRIGHT LAW

Copyrighting Couture: An Examination of Fashion Design Protection and Why the DPPA and IDPPPA are a Step Towards the Solution to Counterfeit Chic 163

CRIMINAL LAW

The Modern History of Probable Cause 377

No Walk in the Dog Park: Drafting Animal Cruelty Statutes to Resolve Double Jeopardy Concerns and Eliminate Unfettered Prosecutorial Discretion 1119

Redeeming an Empty Promise: Procedural Justice, the Crime Victims' Rights Act, and the Victim's Right to be Reasonably Protected from the Accused 47

EMPLOYMENT LAW

Justifying GINA 1051

ENTERTAINMENT LAW

Strictly Business: A Historical Narrative and Commentary on Rock and Roll Business Practices 213

PATENT LAW

In the Wake of Therasense & Nisus Corp., *How Can Patent Attorneys Defend Themselves Against Allegations of Inequitable Conduct?* 487

PROPERTY LAW

Nonuse and Easements: Creating a Pliability Regime of Private Eminent Domain 1

Property Rights, Property Wrongs, and Chattel Disposession Under Self-Storage Leases 1015

SECURITIES LAW

Proceed at Your Peril: Crowdfunding and the Securities Act of 1933 879

TAX LAW

Sales to Grantor Trusts: A Case Study of What the IRS and Congress Can Do to Curb Aggressive Transfer Tax Techniques 973

TORT LAW

Tort Law-Comparative Fault-Original Tortfeasor Rule in Tennessee 259

Why Civil Recourse Theory is Incomplete 431

TABLE OF CASES

A

Abood v. League of Women Voters of Alaska 311, 354
Abou-Sakher v. Humphreys Cnty. 321
ACM P'ship v. Comm'r 1000
Adhi Parasakthi Charitable, Med., Educ. & Cultural Soc'y of N. Am. v. Twp. of W. Pikeland (E.D. Pa. 2010) 535, 555-56, 558-59
AFL-CIO v. Eu 697, 745-47
Ala. Bd. of Trs. V. Garrett 348
Alatech Healthcare, L.L.C. v. United States 70
Albuquerque v. Nor-Am Agric. Prods. 473
Ali v. Fisher 265, 267, 280
Allee v. Medrano 124
Allen Organ Co. v. Kimball Int'l, Inc. 494
Alloway v. Gen. Marine Indus. 473
Am. Bank & Trust Co. v. Cmty. Hosp. of Los Galos-Saratoga, Inc. 477
Am. Sheet & Tin Plate Co. v. Pittsburgh & L.E.R. Co. 1125
Amrein v. State 1136, 1144, 1149, 1157
Andersen v. UNIPAC-NEBHELP 291, 295, 296, 299
Anderson v. Creighton 132, 136
Anderson v. Gus Mayer Boston Store of Del. 1089
Andrews v. S.D. Student Loan Assistance Corp. 293
Animal Fair, Inc. v. AMFESCO Indus. 172
Appalachian Power Co. v. Pub. Serv. Comm'n 316
Applied Materials, Inc. v. Multimextrixs 498-99
Ariel v. Town of Kingston 63, 66-67
Arizona v. California 619, 621, 622, 623, 631
Arizona v. Roberson 575-77
Ark. Gazette Co. v. Pickens 315, 324
Ark. River Packet Co. v. Hobbs 262
Arruda v. Donham & Dover Inv. Props., Inc. 1026
Aschenbach v. Covenant Living Centers-North, Inc. 899
Asgeirsson v. Abbott 314, 333, 334, 339, 353
Ashcroft v. ACLU 340
Ashcroft v. Iqbal 141, 144
Assoc's v. City of Newark 50
Associated Builders & Contractors v. Perry 505, 525
Associated Press v. Bd. of Pub. Educ. 327
Atkinson v. Hemphill 263, 269-70, 272, 273, 278, 282
Austin v. Mich. State Chamber of Commerce 351
Ayala v. Jamaica Sav. Bank 905

B

Baker v. Carr 680-82, 685
Banco Popular de Puerto Rico v. Greenblatt 510-11
Bankers Life & Casualty Co. v. Holland522
Banks v. Elks Club Pride of Tenn. 259-61, 270-86
Banks v. Sallie Mae Servicing Corp. 292, 296, 299, 300
Barker v. Hazetine 746
Baze v. Rees 860, 862, 864-65
Bd. of Cnty. Comm'rs of Boulder Cnty. Colo., v. Rocky Mountain Christian Church (2011) 565
Bd. of Regents of State Colleges v. Ruth 56
Bd. of Trustees v. Bd. of County Comm'rs 322
Becklenberg v. Comm'r 977
Becks v. Emery-Richardson, Inc. 899-900
Beckwith v. Philby 420
Bell v. United States 1155
Benson v. Allphin 132, 133, 134
Berk v. Bates Adver. USA, Inc. 1085

Bervoets v. Harde Ralls Pontiac-Olds, Inc. 261, 264, 265
Bi-Metallic Inv. Co. v. State Bd. of Equalization 326, 327
Bigelow v. Howze 320
Biscan v. Brown 262, 265, 283
Bivens v. Six Unknown Named Agents of Fed. Bureau of Narcotics 108, 141
Blethen Me. Newspapers, Inc. v. State (Me. 2005) 357
Blevins Elec., Inc. v. First Am. Nat'l Bank 294
Blockburger v. United States 1148
Blond v. Stubbs 276
Boddie v. Am. Broad. Co. 127, 133
Boek v. Wong Hong 462
Bogardus v. Comm'r 1001
Bolte v. Home Ins. Co. 518, 527-28
Bond v. Floyd 346-347
Boos v. Barry 334
Boushehry v. State 1136
Bowers v. Hardwick 789
Bowsher v. Synar 503
Boyd v. United States 414
Bragdon v. Abbott 1082, 1084, 1085, 1086-87
Bram v. United States 390
Brandir Int'l, Inc. v. Cascade Pac. Lumber Co. 172
Branzburg v. Hayes 127, 132, 151
Breed v. Jones 1122
Brinegar v. United States 134
Brookwood Area Homeowners Ass'n v. Mun. of Anchorage 315, 319
Browder v. Morris 268
Brown & Williamson Tobacco Corp. v. Fed. Trade Comm'r 327
Brown v. Chadsey 421
Brown v. City of Oneonta 128, 131, 141
Brown v. Comm'r 1003
Brown v. Hartlage 347
Brown v. State 1146
Brown v. Wal-Mart Disc. Cities 264-65.
Brum v. Town of Dartmouth 67
Brunner v. N.Y. State Higher Educ. Servs. Corp. 293
Buck v. Bell 1099
Bultron v. State (Del. 2006) 601-02
Burlington Indus., Inc. v. Davco Corp. 493
Burns v. Erben 388, 420, 422, 423
Burson v. Freeman 337
Bush v. Gore 680, 684, 796, 803
Butler v. Biocore Med. Technologies, Inc. 520

C

Caldwell v. Lambrou 322, 323
Califano v. Boles 349
California v. Greenwood 1109
Campbell v. Superior Court 1121
Camper v. Minor 264
Canadian Tarpoly Co. v. U.S. Int'l Trade Comm'n 512
Capital Info. Grp. v. State (Alaska 1996) 356
Carey v. Brown 332, 335-37, 341-42
Carey v. Piphus 138
Carmen v. Idaho 697, 712, 714
Carol Barnhart, Inc. v. Economy Cover Corp. 173-74
Carroll v. United States 394
Carroll v. Whitney 265, 267, 277, 280, 283
Carter v. City of Nashua 324
Castetter v. Mr. "B" Storage 1027
Chabad of Nova, Inc. v. City of Cooper City (S.D. Fla. 2008) 535, 543-44
Chae v. SLM Corp. 287
Chanel, Inc. v. French 175
Chapman v. Higbee Co. 147
Chavez . Martinez 578, 584
Cherry v. City of New York 513
Chevron U.S.A., Inc. v. Natural Res. Def. Counsel, Inc. 991, 992, 995, 1012
Chevron USA Inc. v. Echazabal 1090
Chiles v. Thornburgh 504
Church of the Hills of Bedminster v. Twp. of Bedminster (D. N.J. 2006) 557
Citizens United v. FEC 352-53
Citizens United v. Fed. Election Comm'n 875
Citron v. Niche Media/Ocean Drive Magazine 1074
City Council v. Cooper 322

City of Alpine v. Abbott 313
City of Boerne v. Flores (1997) 539-40
City of Chicago v. Elevated Props., L.L.C. 1139
City of Cincinnati v. Discovery Network, Inc. 332, 334
City of Cleburne v. Cleburne Living Ctr. 348, 545
City of Erie v. PAP's AM 334
City of Fayetteville v. Edmark 318
City of Los Angeles v. Alameda Books (2002) 560
City of Miami Beach v. Berms 328
City of Naples Airport Auth. v. FAA 504
City of Stilwell, Okla. v. Ozarks Rural Elec. Co-op. 508
Civil Liberties for Urban Believers v. City of Chicago (7th Cir. 2003) 532-33
Clark v. Jeter 349
Clements v. Fashing 351
Comm'r v. Duberstein 1001
Coach Leatherware Co. v. Ann Taylor, Inc. 176
Cochran v. Safeguard Self-Storage, Inc. 1025
Cockerham-Ellerbee v. Town of Jonesville 58, 61
Coggin v. Davey 311, 354
Coggs v. Bernard 1034
Cole v. State (Colo. 1983) 312, 330
Coleman v. Miller 676-78, 680, 697, 713-14, 801-02
Comm'r v. Kann's Estate 997
Comm'r v. Proctor 998-99
Common Cause of Utah v. Utah Pub. Serv. Comm'n 322
Common Cause v. Sterling 322
Commonwealth v. Babb (Mass. 1994) 599
Commonwealth v. Certain Lottery Tickets 387
Commonwealth v. Hyde (Mass. 2001) 127
Commonwealth v. Means (Mass. 2009) 599, 602
Commonwealth v. Melton 1147, 1158
Congregation Koi Ami v. Abington Twp. (E.D. Pa. 2004) 540
Connell v. Town of Hudson 118
Connick v. Myers 339, 344, 345
Conservation Comm'n of Westport v. Beaulieu 346
Consol. Edison Co. v. Pub. Serv. Comm'n 332
Contemporary Indus. Corp. v. Frost 328
Continental Mktg. Corp. v. SEC 886
Cook v. Boorstin 504
Cook v. Gralike (2001) 746
Cook v. Pub. Storage, Inc. 1027, 1047, 1048
Cook v. R.I. Dep't of Mental Health, Retardation, and Hosps. 1089
Cooper v. Aaron 738
Cooter & Gell v. Hartmaex Corp. 488
Cornman v. N.P. Dodge Mgmt. Co. 1086
Corr. Servs. Corp. v. Malesko 144
Cotter v. Mass. Ass'n of Minority Law Enforcement Officers 505
Cottonwood Christian Ctr. v. Cypress Redev. Agency (C.D. Cal. 2002) 540, 556-57
Counselman v. Hitchcock 571
County of Riverside v. McLaughlin 585
Courson v. McMillian 131, 132
Cracraft v. City of St. Louis Park 60-61, 63
Craig v. Boren 349, 842
Crummey v. Comm'r 1007
Culbreath v. Dukakis 510

D

Davis v. Hildebrandt 710
Davis v. Passman 141
Decker v. Campus 117
DeLuca v. Long Island Lighting Co. 512
DeShaney v. Winnebago Cnty. Dep't of Soc. Serv's 53, 56-57, 60
Diamond v. Charles 503, 515
Dickerson v. United States 577, 584
Didrickson v. U.S. Dep't of Interior 505
Dillon v. Gloss 675-76, 681, 712-13, 749
Dixon v. United States 987
Dodge v. Woolsey 743
Doe v. Calumet City 63
Doe v. Linder Constr. Co. 262, 275
Doe v. Santa Fe Indep. Sch. Dist. 327

Donaldson v. United States 506
Donovan v. Priest 746, 804
Doran v. Petroleum Mgmt. Corp. 915
Dorrier v. Dark 321, 328
Dubey v. Pub. Storage, Inc. 1027
Dumas v. State ex rel. Dep't of Culture, Recreation & Tourism 164
Dunn v. Blumstein 348, 350
Durant v. State (Okla. 2008) 116
Duxbury Trucking, Inc. v. Mass. Highway Dep't 142
Dyer v. Blair 697, 703, 711-12, 741, 743, 787

E

Eakins v. Roosevelt 418
Eanes v. State 421
Eastwold v. New Orleans 327
Ebeling v. Morgan 1148
Edwards v. Arizona 570, 574-75, 577
Edwards v. Sisler 263, 283-84
EEOC v. Arabian Am. Oil Co. 1113
EEOC v. Woodbridge Corp. 1077
Elder v. Holloway 132
Elkins v. District of Columbia 137
Elliot Industries v. BP American Production Co. 512
Elliot-Park v. Manglona 57
Emp't Div., Dep't of Human Res. of Or. v. Smith (1990) 538-39
Englewood Citizens for Alternate B v. Town of Englewood 321
Entick v. Carrington 388
Ericsson, Inc. v. Interdigital Comm. Corp. 511, 515, 524-25, 530
Erkenbrecher v. Cox 737
Ernst & Ernst v. Hochfelder 927-28
Escola v. Coca Cola Bottling Co. 472, 473
Espinosa v. United Student Aid Funds, Inc. (9th 2008) 288, 289, 290
Espinosa v. United Student Aid Funds, Inc. (D. Ariz. 2006) 290
Estate of Bongard v. Comm'r 997
Estate of Cristofani v. Comm'r 1001
Estate of Graves v. City of Circleville 58
Estate of Harper v. Comm'r 997
Estate of Hillgren v. Comm'r 997
Estate of Jenkins 270
Estate of McCall v. United States 478, 479
Estate of Schutt v. Comm'r 997
Estate of Stone v. Comm'r 996-97
Estate of Strangi v. Comm'r 997, 1001
Estate of Thompson v. Comm'r 997
Estate of Wal v. Comm'r 988
Eu v. San Francisco Cnty. Democratic Cent. Comm. 347
Evangelatos v. Superior Court 482
Evans Med. Ltd. v. Am. Cyanamid Co. 150
Ex parte Brown 397
Ex parte Bushnell 50
Ex rel Johnson v. South Carolina Dep't of Soc. Serv's 54
Ezell v. Cockerell 61, 63

F

Fashion Originators Guild of Am. v. Fed. Trade Comm'n (2d Cir 1940) 180-81
Fashion Originators Guild of Am. v. Fed. Trade Comm'n (U.S. 1941) 180-81
FEC v. Colo. Republican Fed. Campaign Comm. 331
Feist Publ'ns, Inc. v. Rural Tel. Serv. Co. 170, 198
Fidelity-Phila. Trust Co. v. Smith 977, 996
First Citizens Fed. Sav. & Loan Ass'n v. Worthen Bank & Trust Co. 900
Fisher v. McGirr 416
Fisher v. United States 1001
Flagg Bros. v. Brooks 1016
Flesh v. Bd. of Trs. 327
Florida v. Bostick 128, 129
Foley v. One Hundred & Eighty Bottles of Liquor 416
Forbes v. Wilson Cnty. Emergency Dist. 325
Ford v. Town of Grafton 62, 63-65, 73
Franklin v. Mazda Motor Corp. 483
Frazer v. Dixon Unified Sch. Dist. 315
Freedom Baptist Church v. Middletown (E.D. Pa. 2002) 545
Fromson v. Citiplate, Inc. 512
Frontiero v. Richardson 349

Fund for Animals, Inc. v. Norton 504

G

Gale v. Storti 134
Garcetti v. Ceballos 344-46
Garrett v. Belmont Cnty. Sherriff's Office 54
Garrett v. Estelle 864
Gatlin v. Green 54
Gen. Elec. Co. v. Process Control Co. 268
Gentry Gallery, Inc. v. Berkline Corp. 492
Gideon v. Wainwright 589, 593
Gilchrist v. O'Keefe (2d Cir. 2001) 594, 597-98
Gilligan, Will & Co. v. SEC 924
Globe Newspaper Co. v. Superior Court327
Gonzales v. Raich 813-14, 828
Goodson Todman Enter. v. City of Kingston Common Council 316
Grace Church of Roaring Fork Valley v. Bd. of Cnty. Comm'rs of Pitkin Cnty. (D. Colo. 2010) 535
Grace Fellowship Church of Loudon Cnty., Inc. Lenoir City Beer Bd. 321
Grace United Methodist Church v. City of Cheyenne (10th Cir. 2006) 540
Graham v. Connor 134
Graham v. John Deere Co. of Kansas City 179
Gralike v. Cook (1999) 746
Gravel v. United States 347
Gray v. Ford Motor Co. 268
Great Lakes Higher Educ. Corp. v. Pardee 291, 295, 297, 298, 299, 304
Greenman v. Yuba Power Prods., Inc. 472, 473
Gregg v. Georgia 859
Gregory v. Helvering 999
Grein v. Bd. of Educ. 318
Griggs v. Duke Power Company 1093
Grove Fresh Distribs., Inc. v. Everfresh Juice Co. 327
Grutter v. Bollinger 504
Guru Nanak Sikh Soc'y of Yuba City v. Cnty. of Sutter (E.D. Cal. 2003) 534
Gutwaks v. Am. Airlines 1086

H

Hahnemann Univ. Hosp. v. Edgar 522-23
Haney v. Miami-Dade Cnty. 124
Hardwick v. Doolittle 81-82
Harlow v. Fitzgerald 136
Harris v. New York 574
Harris v. State (Del. 2002) 129, 130
Hawke v. Smith 679, 703, 711, 744-45, 747, 799, 800, 801, 802, 803
Hawley v. Butler 422
Haworth, Inc. v. Steelcase 515
Hayes v. Jackson Parish Sch. Bd. 327
Hays Cnty. Water Planning P'ship v. Hays Cnty. 312, 330, 331, 344
Heffron v. Int'l Soc'y for Krishna Consciousness 342
Helvering v. Clifford 1002-03
Henningsen v. Bloomfield Motors, Inc. 471
Hermes Int'l v. Lederer de Paris Fifth Ave., Inc. 190
Herren v. Bowyer 134, 135
Heyen v. United States 975
Hicks v. State 59
Hill v. Colorado 332, 334, 335-37
Hill York Corp. v. Am. Int'l Franchises, Inc. 914
Hillsboro-West End Neighborhood Ass'n v. Metro Bd. of Zoning Appeals 319
Hinderlider v. La Plata River & Cherry Creek Ditch Co. 618
Hinds Cnty. Bd. of Supervisors v. Common Cause of Miss. 319
Hitt v. Mabry 322
Hocking v. Dubois 893
Hodgson v. United Mine Workers 510
Holeski v. Lawrence 318
Hollingsworth v. Virginia 669-70, 711, 737, 803
Holman v. Comm'r 1001
Home Concrete & Supply, LLC v. United States 994
Houchins v. KQED, Inc. 864
Hunssinger v. Rockford Bus. Credits, Inc. 905
Hutson v. Felder 134

I

Idaho v. Freeman 697, 712, 714
Illinois v. Allen 594
Illinois v. Gates 412
Illinois v. Milwaukee 621
Illinois v. Perkins 576
In re Appeal of the Order Declaring Annexation Dated June 28, 1978 316
In re Cont'l Ill. Sec. Litig. 327
In re Eshostar Communications Corp. 512
In re Galey 305
In re Goldstein 520
In re Hanson 292, 296, 300
In re Huff 416
In re Initiative Petition No. 364 803
In re Iowa Freedom of Info. Council 327
In re Mammel 296
In re Opinion of the Justices (Ala. 1933) 745
In re Opinion of the Justices (Me. 1933) 706-07, 711, 745
In re Opinion of the Justices (N.C. 1933) 703, 711, 732
In re Special Investigations No. 228 402
In re Stevens 296
In re Washington Post Co. 327
In re Washington Pub. Power Supply Sys. Sec. Litig. 908
In re Williams 516, 518
In re Williams Sec. Litig.-WCG Subclass 934
Int'l Bhd. Of Teamsters v. Daniel 899
Int'l Union v. Scofield 511, 527
Internal Rev. Serv. v. Taylor 305

J

J Choo (Jersey) Ltd. v. Towerstone Ltd. 193
Jackson v. Hamilton 269, 270
Jade Trading, LLC v. United States 1000
Javins v. First Nat'l Realty Corp. 1041
Jean W. v. Commonwealth 62
Jenness v. Fortson 351
Jerman v. Carlisle, McNellie, Rini, Kramer, & Ulrich, LPA 322
Johnson v. Zerbst 575, 579, 582, 593
Johnston v. Metro. Gov't of Nashville & Davidson Cnty. 320-21
Jones v. Inter-County Imaging Centers 1089
Jones v. Phyfer 54
Jones v. State (Del. 1999) 131
Jordan v. Mich. Conference of Teamsters Welfare Fund 508, 510

K

Kaho'Ohanohano v. Dep't of Human Servs. 58
Kan. City Star Co. v. Shields 318
Kane v. U-Haul Int'l Inc. 1041
Kane v. U-Haul Int'l, Inc. 1024
Karcher v. May 516
Karen Millen Ltd. v. Dunnes Stores 193, 203
Katskee v. Blue Cross/Blue Shield of Neb. 1087
Kawananakoa v. Polyblank 59
Keith v. Rizzuto 70
Kenna v. U.S. Dist. Court 80
Kentucky v. Indiana 619
Kerr v. U.S. Dist. Court 523
Kieselstein-Cord v. Accessories by Pearl, Inc. 172-73
Kimbell v. United States 997
Kimble v. Swackhamer 697, 711, 746-47
Kimel v. Fla. Bd. of Regents 348
King v. Ambs 136
King v. Superior Court (Cal. Ct. App. 2003) 600-01
Kingsdown Med. Consultants, Ltd. v. Hollister Inc. 493, 498, 523
Knight v. Iowa Dist. Ct. of Story Cnty. 328
Knitwaves, Inc. v. Lollytogs Ltd. 176, 178
Konikov v. Orange Cnty., Florida (M.D. Fla. 2004) 535, 545-46, 554, 564
Kootenai Trive of Idaho v. Veneman 505, 509, 513, 525, 527, 530
Korczak v. Sedeman 506
KQED, Inc. v. Vasquez 864
Kramer v. Union Free Sch. Dist. 350

L

LaDuke v. Nelson (1982) 124
LaDuke v. Nelson (1985) 123, 150, 151
Lakewood, Oh. Congregation of Jehovah's Witnesses, Inc. v. City of Lakewood (6th Cir. 1983) 537
Landrau v. Solis Betancourt 118, 120
Lathers v. U-Haul Co. 1024
Law & Info. Servs., Inc. v. Riviera Beach 327
Lawrence v. City of Cambridge 63-64
Lawrence v. Texas 349, 789, 814
League of Women Voters of Me. v. Gwadosky 746, 804
Leicester v. Warner Bros. 119, 120
Leichtamer v. Am. Motors Cop. 473-74,
LeMay v. U.S. Postal Serv. 70
Leone v. City of Chicago 60-61
Leser v. Garnett 681, 703, 711, 745, 801-02
Limbaugh v. Coffee Med. Ctr. 267-68
Lincoln Nat'l Bank v. Herber 905
Linda R.S. v. Richard D. 57, 81
Lively v. Hirschfeld 914
Lloyd v. Comm'r 997-98
Lochner v. New York 789
Lois Sportswear v. Levis Strauss & Co. 190
Long v. Educ. Credit Mgmt. Corp 293, 294
Louis Vuitton Malletier v. Dooney & Burke, Inc. (2d Cir. 2006) 175-76
Louis Vuitton Malletier v. Dooney & Burke, Inc. (S.D.N.Y 2007) 176
Louisiana v. Texas 618
Low v. Penn Self Storage 1045
Lowrey v. Gridley 386, 415
Lugar v. Edmondson Oil Co. 147
Lujan v. Defenders of Wildlife 502, 506
Lusardi v. Lechner 522, 529

M

Madison Joint Sch. Dist. V. Wis. Emp't Relations Comm'n 326, 327
Madison v. Riter (4th. Cir. 2003) 538
Mahoney v. Nat'l Org. for Women 141
Malley v. Briggs 136
Marbury v. Madison 789, 799
Marshall Field & Co. v. Clark 810
Martin v. T.V. Tempo, Inc. 900
Maryland v. Louisiana 622, 629, 631
Maryland v. Shatzer (2010) 569-71, 579-80, 582-85
Maryland v. Shatzer (Md. 2008) 570
Massachusetts v. Oakes 133
Matthews v. Comm'r 1003
Matthews v. Pickett County 66
Mattox v. City of Forest Park 142
Maurice River Twp. Bd. of Educ. V. Maurice River Twp. Teachers' Ass'n 328
Mayhew v. Wilder 311, 354-55
Mayo Found. For Med. Educ. & Research v. United States 991, 992, 994
Mayor & City Council v. El Dorado Broad Co. 320
Mazer v. Stein 163, 171-72, 187
MCA Records, Inc. v. Newton-John 251
McAlister v. Methodist Hosp. of Memphis 263
McComas v. Bd. of Educ. 312, 329, 358
McConnell v. FEC 331, 503-05, 524-25
McCulloch v. Maryland 651, 705, 805, 868
McElroy v. Strickland, Knox Cnty. Ch. 320
McIntyre v. Balentine 264, 265
McNeil v. Wisconsin 574, 576, 580
Mead v. Warner Pruyn Div., Finch Pruyn Sales, Inc. 475
Med. Liab. Mut. Ins. Co. v. Alan Curtis LLC 504
Memphis Cmty. Sch. Dist. v. Stachura 138
Mercer v. Vanderbilt Univ., Inc. 265, 268-69, 280, 281-83
Mersmann v. United Student Aid Funds, Inc. 292, 294, 295, 296, 298, 299, 305
Metro-Goldwyn-Mayer Studios, Inc. v. Grokster, Ltd. 208
Metro. Air Research Testing Auth., Inc. v. Metro. Gov't of Nashville 321
Metromedia, Inc. v. San Diego 50
Miami-Dade Cnty. v. Miller 60
Michigan v. Summers 135
Michigan v. Tucker 574
Miller v. California 133
Miller v. Cen. Chinchilla Grp., Inc. 886

Miller v. Moore 746
Mills v. State 1138, 1144
Milton v. Srivner, Inc. 1081-82
Minn. State Bd. v. Knight 326, 327
Minnick v. Mississippi 577
Miranda v. Arizona 572, 575, 577
Mississippi v. Louisiana 618
Mitchell v. Forsyth 132
Moberg v. Indep. Sch. Dist. 340
Molins PLC v. Tetron, Inc. 493-94, 498
Montefusco v. Nassau Cnty. 125, 126, 133, 139
Moody v. Gulf Ref. Co. 262
Moore v. City of Chi. Heights 57
Moore v. Regents of Univ. of Cal. 1108-09
Moreland v. Dep't of Corp. 900
Morgan v. Kroupa 1137
Morrison v. Nat'l Australia Bank, Ltd. 892
Mott v. Mott 418
Mullane v. Central Hanover Bank & Trust Co. 299
Murphy v. Edmonds 482, 483
Myers v. McGrady 60

N

N.J. Soc. for Prevention of Cruelty to Animals v. N.J. Dept. of Agric. 1134
N.Y. Times v. Sullivan 339
NAACP v. Alabama 353
Naifeh v. Valley Forge Life Ins. Co. 262m 285
Nat'l Cable & Telecomms. Ass'n v. Brand X Internet Servs. 995, 996
Nebraska v. Wyoming 621
New Jersey v. New York 620-21, 631
New Jersey v. TLO 428
New York v. Quarles 575-76
New York v. United States 820
Newman v. Graddick 327
Nisus Corp. v. Perma-Chink Sys., Inc. (E.D. Tenn. 2006) 494-95, 513, 530
Nisus Corp. v. Perma-Chink Sys., Inc. (Fed. Cir. 2007) 507, 513-14, 516, 519, 521-22, 524-30
NLRB v. Sears, Roebuck & Co. 361
Nobleman v. Am. Sav. Bank 287
North Dakota v. Minnesota 619

O

O'Connor v. Ortega 1110
Oaks v. Connors 483
Ohio v. Lapping 1149, 1150
Ohio v. Wyandotte Chems. Corp. 618
Okin v. Village of Cornwall-on-Hudson Police Dep't 54-55
Oklahoma v. Texas 620, 629, 631
Opinion of the Justices to the Senate (1977) 703, 732
Orr v. Wal-Mart Stores, Inc. 1091
Ortega Melendres v. Arpaio 129, 135
Oser v. Wal-Mart Stores 475
Owens v. Truckstops of Am. 261, 264, 266-67, 277, 279-83

P

Palsgraf v. Long Island R.R. Co. 435, 456
Parole & Prob. Comm'n v. Thomas 318
Partain v. Oconee Cnty. 60
Paul v. Virginia 50
Payton v. New York 390
Payton v. Rush-Presbyterian-St. Luke's Med. Ctr. 147
Pena v. Deprisco 54-55
Penhouse Int'l, Ltd. v. Playboy Enters. 512-13
People ex rel. Cross v. Martin 418
People ex rel. Difanis v. Barr 312
People ex rel. Doherty v. Police Com'rs of New York 418
People v. Baniqued 1127
People v. Berberrich 416
People v. Bowie 1147
People v. Counts 1152
People v. Davis 1150-51
People v. Dieguez 1151
People v. Harris 1147
People v. Jakira 416
People v. Rae 1151
People v. Rouser 1147
People v. Sanchez 1149, 1150, 1151, 1157
People v. Stankewitz 1151

People v. Storm 580
People v. Toynbee 416
People v. Wakeford 1157
People v. Zavala 1150
Percision Instrument Mfg. Co. v. Hartford Co. 493
Percision Specialty Metals, Inc. v. United States 517, 519
Perry v. United States 1003
Peschel v. City of Missoula 58
Peter Pan Fabrics, Inc. v. Brenda Fabrics, Inc. 174, 200
Petter v. Comm'r 998-99
Phoenix Ins. Co. v. Erie Transp. Co. 453
Pickering v. Bd. of Educ. 344, 345
Pile v. Brandenburg 58
Poland v. Educ. Credit Mgm. Corp. 298
Police Dep't of Chi. v. Mosley 351
Pollock v. Farmers' Loan & Trust Co. 815
Portland Audubon Soc'y v. Hodel 504, 525
Powell v. McCormick 697, 738
Pratt v. Hill 421
Press Enter. Co v. Super. Ct. of Cal., 478 U.S. 1 (1986) 326
Press Enter. Co. v. Super. Ct. of Cal., 464 U.S. 501 (1984) 326-27
Price-Cornelison v. Brooks 57
Printz v. United States 820
Prior v. Norland 703, 711
Publicker Indus. V. Cohen 327
Pulitzer Publ'g Co. v. McNeal 322

Q

Qualitex Co. v. Jacobson Prods. Co. 177
Qualls v. Lack's Stores, Inc. 1086

R

R.A.V. v. City of St. Paul 333, 341
R.M.S. Titanic, Inc. v. Haver 118, 119, 120, 146, 151
Raas v. State 60
Radke v. County of Freeborn 58
Randall v. Sorrell 331
Rangra v. Brown, 566 F.3d 515 (5th Cir. 2009) 339, 345, 346, 347
Rangra v. Brown, 576 F.3d 531 (5th Cir. 2009) 313
Rangra v. Brown, 584 F.3d 206 (5th Cir. 2009) (en banc) 313
Rauenhorst v. Comm'r 987-88
Relford v. Lexington-Fayette Urban Cnty. Gov't 128
Reno v. ACLU 331
Renton v. Playtime Theatres, Inc. 333
Republican Party of Minn. v. White 331-32, 338-39, 346
Resolution Trust Corp. v. Block 267, 268, 281
Revak v. SEC Realty Corp. 889-90
Revell v. McCaughan 261, 263
Reves v. Ernst & Young 890, 900, 906
Reynolds v. Sims 350, 666-67, 690
Rhode Island v. Palmer 737
Richmond Newspapers, Inc. v. Virginia 127, 327
Risbridger v. Connelly 136
Rivera v. United States 479-80
Roberson v. Copeland 321
Roberts v. United States 738
Robinson v. Att'y Gen. of U.S. 357
Roche v. Evaporated Milk Assn. 522
Rocky Mountain Christian Church v. Bd. of County Comm'rs of Boulder Cnty. (D. Colo. 2009) 533, 540, 547, 549, 551, 554, 557, 564
Rocky Mountain Christian Church v. Bd. of County Comm'rs of Boulder Cnty. (D. Colo. 2010) 534, 535-36, 546-54, 556, 557, 558, 560-61, 566-67
Roeder v. Islamic Republic of Iran 504
Rohan v. Swain 421
Romanski v. Detriot Entm't, L.L.C. 147, 148
Rothstein v. United States 976, 988, 989
Ruehle v. Educ. Credit Mgmt. Corp. 292, 296, 300, 301
Ruiz v. Estelle 504-05
Russell v. Shuster 421

S

Saenz v. Roe 350

Salyer Land Co. v. Tulare Lake Basin Water Storage Dist. 350
Samuelson v. McMurtry 280
San Antonio Indep. Sch. Dist. v. Rodriguez 348
San Jose Christian Coll. v. City of Morgan Hill (9th Cir. 2004) 533
San Juan Cnty. v. United States (10th Cir. 2007) 505, 507, 510
San Juan Cnty., Utah v. United States (10th Cir. 2005) 503
Sanders v. City of Fort Smith 322
Sandoval v. Bd. of Regents of Univ. 312, 330, 331
Santor v. A & M Karagheusian 473
Sarkes Tarzian, Inc. v. Legislature of the State of Nev. 311, 354
Satterfield v. Breeding Insulation Co. 262
Savard v. Rhode Island 145
Saxbe v. Wash. Post Co. 864
Scarves by Vera, Inc. v. United Merchs. & Mfrs. 174
Schad v. Borough of Mount Ephraim (1981) 543
Scheier v. City of Snohomish 135, 137
Schofield v. First Commodity Corp. 887
Schwertfager v. Boynton Beach 1086
Sciolino v. Ryan 320
Seattle-First Nat'l Bank v. Volkswagen of Am., Inc. 475
SEC v. Allison 926
SEC v. C.M. Joiner Leasing Corp. 895, 905, 906
SEC v. Chinese Consol. Benevolent Ass'n 925
SEC v. Datronics Eng'rs, Inc. 923
SEC v. Edwards 886, 889, 893, 895-96, 923
SEC v. Glenn W. Turner Enters. 888, 889
SEC v. Koscot Interplanetary, Inc. 888, 889, 891
SEC v. Life Partners, Inc. 923
SEC v. Murphy 914
SEC v. Phan 926
SEC v. Ralston Purina Co. 913, 914
SEC v. U.S. Realty & Improvement Co. 509
SEC v. W.J. Howey Co. 886, 887-88, 889, 892, 893, 895, 896, 897-98, 903, 923
Sec. Self-Storage v. Pauling 1027
Selle v. Gibb 205
Serviss v. State, Dep't of Natural Res. 63
Seymour v. Hug 518
Sharondale Constr. Co. v. Metro. Knoxville Airport Auth. 321
Shaw v. Hunt 504, 509
Shelton v. Comm'r 989-90
Shepard v. Bradford 60
Sherbert v. Verner (1963) 538
Sherer v. Linginfelter 267
Siegman v. District of Columbia 117
Silco Inc. v. Comm'r 987-88
Silverthorne Lumber Co. v. United States 412
Simon v. E. Kentucky Welfare Rights Org. 502
Simpson v. United States 995
Skemp v. Comm'r 1003
Skidmore v. Swift & Co. 990
Smiley v. Citibank 990
Smiley v. Holm 678-79, 710
Smith v. Kenny 125, 131
Smith v. Mo. Pac. R.R. 511
Smith v. Pa. Emps. Benefit Trust Fund 312, 330
Smith v. Union Bank 739
Smith v. Zoning Bd. of Appeals of Town of Greenwich (Conn. 1993) 561
Sofie v. Fibreboard Corp. 482
South Carolina v. North Carolina (2007) 614
South Carolina v. North Carolina, (2009) 614-16, 621-29, 630-31, 632-34
South Dakota v. Dole 815
South v. Maryland 59-60
Spagnola v. Mathis 145
St. Cloud Newspapers, Inc. v. Dist. 742 Cmty. Sch. 312, 315-16, 330
St. Louis & S.F.Ry. Co. v. Mathews 50
Stallworth v. Monsanto Co. 510
Standard Heating & Air Conditioning Co. v. City of Minneapolis 504
Stanfield v. Neblett 279
Star Scientific, Inc. v. R.J. Reynolds Tobacco Co. 494
State ex rel Donnelly v. Myers 703, 711

State ex rel Harper v. Waltermire 745-46
State ex rel Tate v. Sevier 711
State ex rel. Akin v. Town of Kingston Springs 321
State ex rel. Gregan v. Koczur 1131
State ex rel. Matthews v. Shelby Cnty. Bd. of Comm'rs 321
State ex rel. Murray v. Palmgren 318, 331, 355-56, 366
State ex rel. Newspapers, Inc. v. Showers 316
State v. Angus 1121, 1149, 1150
State v. Bauer 1120
State v. Blom 1137
State v. Boykin (S.C. Ct. App. 1996) 599-600
State v. Burkhart (Tenn. 2001) 322
State v. Carruthers (Tenn. 2000) 592-93, 603, 604, 605-07, 609-10
State v. Claiborne 1127
State v. Conger 81, 82
State v. Eldredge 1137, 1144
State v. Farnsworth 1137, 1161
State v. Holmes (Tenn. 2010) 589, 590, 591, 592, 605-12
State v. Holmes (Tenn. Crim. App. 2007) 589, 590
State v. Holmes (Tenn. Crim. App. 2009) 591, 592
State v. Horton 1121
State v. Kinney 416
State v. Lapping 1149, 1157
State v. Lashinsky (N.J. 1979) 117, 118
State v. MacFarlane 1137
State v. Malusky 1121
State v. McCann 411
State v. Montgomery (N.C. Ct. App. 2000) 600
State v. Myers 1146-47
State v. Northington (Tenn. 1984) 589, 593
State v. Patton (Okla. 1992) 318
State v. Shafer 1147-48
State v. Small (Tenn. Crim. App. 2009) 603-04
State v. Spiritous Liquors (1852) 411
State v. Spiritous Liquors (1855) 411
State v. Staples 411
State v. Taylor (N.J. 1955) 124, 151
State v. Twenty-Five Packages of Liquor 416
State v. Webb 1121, 1138, 1144
State v. Williams 1163
State v. Willis (Tenn. Crim. App. 2009) 604
State v. Wilson 1137
State v. Yorczyk 1139, 1144
State v. Ziemann 1139
Stearns v. Titus 428
Stewart v. Abraham 124
Stone v. Dana 389
Strauss v. Belle Realty Co. 456
Sullivan v. Comm. on Admission and Grievances 519
Summers v. Harris Constr. 63
Sutter Bay Assocs. v. Cnty. of Sutter 340, 358
Sutton v. United Air Lines, Inc. 1079, 1086
Swallows Holding, Ltd. v. Comm'r 992, 993, 994

T

Tarasoff v. Regents of the Univ. of California 435
Teaford v. Ford Motor Co. 516
Teague v. Bakker 898
Tenn. Student Assistance Corp. v. Hood 294
Tenney v. Brandhove 347
Terry v. Ohio 128
Texas v. Johnson 335
Texas v. Louisiana 621, 629, 631
Texas v. White 351
Therasense, Inc. v. Becton, Dickinson & Co. (Fed. Cir. 2010) 495, 489, 498, 528
Therasense, Inc. v. Becton, Dickinson & Co. (N.D. Cal. 2008) 488, 496, 524
Thomas v. City of Cleveland 138
Thomas v. Cnty. of L.A. 124
Thompson v. Lantz 123
Timmreck v. Munn 887
Torres v. Oakland Scavenger Co. 512
Town of Castle Rock v. Gonzales 55-56, 60, 73

TrafFix Devices, Inc. v. Mktg. Displays, Inc. 178
Transps., Inc. v. Perry 261, 262, 263, 283, 285
Travelers Indem. Co. v. Bailey 302
Trbovich v. United Mine Workers 507-08, 510
Troy v. Herndon 269, 270
Turner Broad. Sys. v. FCC 331, 332, 333, 342
Tutor-Saliba Corp. v. Comm'r 991

U

U.S. Postal Serv. v. Brennan 504
U.S. Term Limits, Inc. v. Hill 849
U.S. Term Limits, Inc. v. Thornton 660, 849-50
Uhler v. AFL-CIO 745
Uhler v. Am. Fed. Of Labor-Cong. Of Indus. Org. 799, 802-03
Underhill v. Royal 900
Union Pac. Res. Co. v. Chesapeake Energy Corp. 493
United Am. Bank v. Gunter 891
United Hous. Found., Inc. v. Forman 886, 889, 895-96, 905
United States v. Ansaldi 1145
United States v. Apollo Energies, Inc. 1144
United States v. Arvizu 129
United States v. Atkinson 1144
United States v. Ballin 810
United States v. Batchelder 81
United States v. Bernal 1129, 1144
United States v. Blake 80
United States v. Butler 815
United States v. Carolene Prods. Co. 875
United States v. Cleveland Indians Baseball Co. 992
United States v. Degendardt 70, 71
United States v. Dickerson 393
United States v. Fowler (5th Cir. 1979) 594
United States v. Goldberg (3d Cir. 1995) 594, 596-97
United States v. Gray 128, 131
United States v. Hansen 1130
United States v. Heaton 97
United States v. Henry 125
United States v. Jennings (3d Cir. 1995) 595,
United States v. Jennings (M.D. Pa. 1994) 595, 596, 600
United States v. Jorn 1122
United States v. Lee 128
United States v. Leggett (3d Cir. 1998) 597-98, 604,
United States v. Leon 412
United States v. Lopez 826-28
United States v. McClain 129, 130
United States v. McLeod (11th Cir. 1995) 594, 595, 600
United States v. Mead Corp. 993
United States v. Nevada 621
United States v. Nixon 326, 356, 360
United States v. O'Brien 331, 332
United States v. Patane 578, 584
United States v. Patkar 97
United States v. Playboy Entm't Grp., Inc. 340
United States v. Rostenkowski 778
United States v. Rubin 50, 70, 71, 97
United States v. Slaerno 50
United States v. South-Eastern Underwriters Ass'n 50
United States v. Sprague 703, 706, 711, 714
United States v. Stevens 1121, 1122, 1124, 1128, 1166
United States v. Talao 518, 525
United States v. Thomas (3d Cir. 2004) 598-99
United States v. Turner 50, 70, 71, 96, 99-100, 102
United States v. Virginia 349, 842
United States v. Zuni 102
United Student Aid Funds, Inc. v. Espinosa (2010) 287, 288, 289, 290, 292, 301, 302, 303, 304, 305, 306
Univ. of Tenn. Arboretum Soc., Inc. v. City of Oak Ridge 321
Utah Ass'n of Counties v. Clinton 507

V

Va. State Bd. of Pharmacy v. Va. Citizens Consumer Council, Inc. 326
Valley Forge Christian Coll. v. Ams. United for Sparation of Chuch & State, Inc. 502
Van Hooser v. Warren Cnty. Bd. of Educ. 324
Vandermark v. Ford Motor Co. 474
Venner v. State 1109
Village of Willowbrook v. Oleck 348
Vision Church, United Methodist v. Vill. of Long Grove (7th. Cir. 2006) 535, 546, 552, 558-59
Volz v. Ledes 261, 264, 265

W

Wal-Mart Stores v. Samara Bros. 176-77, 178
Walker v. City of Mesquite 519-20
Wallace v. Williams 419
Walton v. Comm'r 984, 1012
Ward v. Rock Against Racism 332, 342
Ware v. James City Cnty., Virginia 137
Warfield v. Alaniz 891
Washington v. Glucksberg 814
Washpon v. Parr 136
Waste Mgmt., Inc. of Tenn. v. Solid Waste Region Bd. of Metro. Gov't 321
Waters v. Churchill 347
Weissman v. Quail Lodge, Inc. 518
Wesberry v. Sanders 666
West Coast Hotel v. Parrish 789
Westfall v. State 1136, 1144
Wexford Cnty. Prosecuting Attorney v. Pranger 318
Wheeler v. O'Connell 468
Whelton v. Educ. Credit Mgmt. Corp. 293, 305
Whimsicality, Inc. v. Rubie's Costume Co. 172
Whipper v. McLendon Movers, Inc. 1026
Whitehead v. Toyota Motor Corp. 265-66
Wilburn v. City of Memphis 279
Wilkie v. Robbins 142, 145
Will v. United States 1122
William v. Rhodes 350-51
Williams v. Eggleston 352
Williams v. United States 528
Williams v. Walker-Thomas Furniture Co. 249-50
Wilson v. Brown 356
Wisconsin v. Yoder (1972) 538-39
Wolfson v. State (Fla. 1977) 315
Worthington v. Kenkel 123, 124
Wynehamer v. People 417
Wyoming v. Colorado 619

Z

Zylka v. City of Crystal (Minn. 1969) 547