Tennessee Law Review

Volume 73
Issue 1 *2005-2006*

Article 1

2006

Volume 73

TENNESSEE LAW REVIEW

Member of the National Conference of Law Reviews

Volume 73 | Fall 2005 | Number 1

CONTENTS

TENNESSEE LAW REVIEW

Member of the National Conference of Law Reviews

Volume 73 Winter 2006 Number 2

CONTENTS

TENNESSEE LAW REVIEW

Member of the National Conference of Law Reviews

Volume 73 | Spring 2006 | Number 3

CONTENTS

TENNESSEE LAW REVIEW

Member of the National Conference of Law Reviews

Volume 73 Summer 2006 Number 4

CONTENTS

SECURING A RELIABLE ELECTRICITY GRID: A NEW ERA IN TRANSMISSION SITING REGULATION?

STEVEN J. EAGLE*

ABSTRACT

A capacious and reliable electric transmission grid is vital to public health and safety, the economy, and national security. Electric production's recent shift from local generation and use of electric power to regional and national markets requires substantially augmented transmission facilities. New transmission capacity, however, has often been blocked by local favoritism, not-in-my-backyard concerns, and by legislative and judicial doctrines mandating approval processes that disregard any benefits that would inure to other markets or states. The Energy Policy Act of 2005[1] contains siting provisions that offer some promise of relief. This article analyzes present conditions and the extent to which the new Act is apt to be efficacious.

I. THE NEED TO SITE A SECURE AND RELIABLE ELECTRICITY GRID

Electricity is of immense importance in our economy and culture.[2] It has no close substitute in many of its functions and is uneconomical to store in anticipation of peak needs, future requirements, or times of crisis.[3] As the recent catastrophic results of Hurricane Katrina indicate, our emergency management system, energy production, and public health are dependent upon the transmission of electricity.[4]

The United States now is undergoing a transition from local command-and-control electric production and distribution to regional market-controlled

* Professor of Law, George Mason University School of Law, Arlington, Virginia (seagle@gmu.edu). The author wishes to thank Cliff Hyra, a student at George Mason Law School, for his excellent research assistance. The author also acknowledges with appreciation research support received from the Critical Infrastructure Protection Project of the National Center for Technology & Law, George Mason University School of Law.

1. Energy Policy Act of 2005, Pub. L. No. 109-58, 119 Stat. 594 (codified in scattered sections of U.S.C.) [hereinafter "2005 Act"].

2. *See* Lois R. Lupica, *Transition Losses in the Electric Power Market: A Challenge to the Premises Underlying the Arguments for Compensation*, 52 RUTGERS L. REV. 649, 652 n.6 (2000).

3. *See id.*

4. *See, e.g.*, Jad Mouawad, *No Quick Fix for Gulf Oil Operations*, N.Y. TIMES, Aug. 31, 2005, at C1 (noting that a large petroleum pipeline company serving the Mid-Atlantic states and New England "warned that its operations were shut down because of 'widespread damage to the electric grid'").

production and distribution. This profound transformation requires changes in federal and state regulatory regimes to ensure the availability of an adequate and reliable supply of electricity throughout the nation. In an age of terrorism, enhanced protection for critical electric infrastructure is also required. Intermingled with these developments is the need for more electric generating stations and distribution lines and towers. This need both creates and is stymied by not-in-my-back-yard ("NIMBY") protests.[5]

The importance and particular nature of electricity does not render its production and distribution immune from basic problems affecting the energy industry in general:

> While much needed energy infrastructure has been added in certain regions in the past several years, particularly with respect to electric generating capacity and natural gas pipelines, critical needs remain and there are numerous examples of abandoned projects, difficult and time-consuming infrastructure siting processes, and strong political opposition. These siting difficulties stem from the friction between a public that is increasingly unwilling to accept the construction of energy and other industrial infrastructure in their local communities, and the growing need to add critical energy infrastructure to meet business and consumer demands and minimize the potential rippling effect on other economic sectors of major infrastructure-related outages or losses.[6]

The transition to market-based electric generation and distribution has many costs. Some involve transitional losses to traditional utilities, as costly assets acquired under expectations (or implicit promises) of adequate cost recovery in a regulated pricing regime may lose substantial value in the new era of market pricing.[7]

"Perhaps the greatest obstacle to the construction of new [electric] transmission [capability] . . . is the age-old problem of gaining approval for new transmission lines."[8]

> Transmission lines, substations, circuit breakers, capacitors, and other equipment provide more than just a highway to deliver energy and power from generating units to distribution systems. Transmission systems both complement and substitute for generation. Transmission generally enhances

---

5. *See* discussion *infra* Part II.D.1.

6. PAUL J. HIBBARD, ANALYSIS GROUP, INC., U.S. ENERGY INFRASTRUCTURE: DEMAND, SUPPLY AND FACILITY SITING 3 (2004), *available at* http://www.energycommission.org/files/finalReport/V.1.a%20-%20US%20Energy%20Infrastructure.pdf

7. For an extended analysis and argument that the government has a constitutional duty to compensate utilities for "stranded costs" resulting from rate deregulation, see generally J. GREGORY SIDAK & DANIEL F. SPULBER, DEREGULATORY TAKINGS AND THE REGULATORY CONTRACT: THE COMPETITIVE TRANSFORMATION OF NETWORK INDUSTRIES IN THE UNITED STATES (1997).

8. ERIC HIRST, EXPANDING U.S. TRANSMISSION CAPACITY 11 (2000), *available at* http://www.eei.org/industry_issues/energy_infrastructure/transmission/hirst2.pdf.

> reliability; lowers the cost of electricity delivered to consumers; limits the ability of generators to exercise market power; and provides flexibility to protect against uncertainties about future fuel prices, load growth, generator construction, and other factors affecting the electric system.
> Because most of the U.S. transmission grid was constructed by vertically integrated utilities before the 1990s, these legacy systems support only limited amounts of inter-regional power flows and transactions. Thus, existing systems cannot fully support all of society's goals for a modern electric-power system.[9]

A. The Restructuring of the Electricity Market

Experts began calling for a change in the structure of the electricity market from its natural monopoly roots in the mid-1980s.[10] They cited studies performed between 1948 and 1974 that documented serious problems with the performance and efficiency of the market, including ineffective regulation and the failure of utilities to coordinate activities and to take advantage of available economies of scale.[11] These issues were estimated to cost the United States economy billions of dollars each year.[12] Consumers and regulators began to take notice of the problem when the price of electricity surged between 1974 and 1984.[13]

With the development of efficient small-scale electricity generation and long-distance transmission, electricity generation was ripe for deregulation and competition, although transmission and distribution remained natural monopolies.[14] "In a restructured market, hundreds of [independent] generators c[ould] compete with each other to generate and sell electricity" for the lowest price.[15] In addition, the dramatic improvement in transmission technology and interconnectedness expanded the appropriate scale on which to regulate the market from a local to a regional, national, or even continental area.[16] The success of restructuring in other industries convinced many that it was time to make changes to the electricity market.[17]

States, Congress, and the Federal Energy Regulatory Commission

---

9. ERIC HIRST, U.S. TRANSMISSION CAPACITY: PRESENT STATUS AND FUTURE PROSPECTS v (2004), *available at* http://www.eei.org/industry_issues/energy_infrastructure/transmission/USTransCapacity10-18-04.pdf.

10. Richard J. Pierce, Jr., *Completing the Process of Restructuring the Electricity Market*, 40 WAKE FOREST L. REV. 451, 453-55, 458 (2005).

11. *Id.* at 453.

12. *Id.*

13. *Id.* at 454.

14. *Id.* at 461-62; Michael Coyn Mateer, Note, *When The Lights Go Out: The Impact of House Bill 6 on Regional Transmission Organizations and the Reliability of the Power Grid*, 12 GEO. MASON L. REV. 775, 809-10 (2004).

15. Pierce, *supra* note 10, at 462.

16. *Id.*

17. *Id.* at 463-64.

("FERC") have been attempting to restructure the electricity market to varying degrees for the past two decades.[18] They have focused on encouraging the growth of independent merchant generators, which use transmission capacity built by others to sell the electricity they produce on the market, and on creating integrated regional electricity markets.[19] In the late 1990's, for example, the FERC issued orders that forced utilities to separate their transmission and generation functions, charge all generators the same rates for transmission access, share transmission information with all generators, and join Regional Transmission Organizations ("RTOs") "or justify to the FERC their reasons for not joining one."[20]

The restructuring effort's success has varied dramatically from state to state. Those states and regions with low-cost electricity, such as the Southeast, Northwest, and lower Midwest, have largely resisted restructuring, fearing that change would increase their prices and rob them of a competitive advantage.[21]

Only the Mid-Atlantic region, New England, New York, and Texas have seen significant success in restructuring.[22] In each of these areas, power is provided mostly by independent merchant generators.[23] Ownership of transmission, distribution, and generation assets is largely disaggregated, and a single regional entity operates the electricity grid.[24] Restructuring has saved the Mid-Atlantic region several billion dollars a year.[25] Elsewhere in the country, however, entrenched opposition groups have effectively blunted the restructuring effort.[26]

Attempts at electricity market restructuring have faced many challenges in addition to regional opposition from market and political actors. Restructuring is best implemented by a federal agency with broad regulatory power, but states have most of the regulatory authority in the electricity industry and the power of the FERC is severely limited.[27] Therefore, it is impossible for restructuring to proceed in a coordinated fashion on a national level.

An even more serious problem is the dramatic shortfall in transmission capacity throughout the nation.[28] This shortage "will eventually doom all restructuring efforts and . . . will yield disastrous results for the entire U.S.

---

18. *See id.* at 468-95.

19. *See* Pierce, *supra* note 10, at 468-95.

20. Mateer, *supra* note 14, at 791-99.

21. Pierce, *supra* note 10, at 459-60.

22. *Id.* at 468-79.

23. *Id.* at 469-70.

24. *Id.*

25. RONALD J. SUTHERLAND, CTR. FOR THE ADVANCEMENT OF ENERGY MKTS., ESTIMATING THE BENEFITS OF RESTRUCTURING ELECTRICITY MARKETS: AN APPLICATION TO THE PJM REGION 5 (2003), *available at* http://der.lbl.gov/pubs/BenefitsOct1Final.pdf.

26. *See, e.g.*, Pierce, *supra* note 10, at 477-79 (detailing the efforts of the "just say no" group in slowing the FERC's attempts to mandate a country-wide standard market design).

27. *Id.* at 466.

28. *Id.* at 469.

market no matter how it is structured, unless and until it is solved."[29] The problem stems, in part, from the lack of FERC jurisdiction over electric transmission line siting.[30]

In addition, some of the FERC's regulatory changes have actually exacerbated the transmission shortage.[31] Order 888,[32] for example, required that any utility owning transmission assets charge the same rate to every company using its transmission lines.[33] Unfortunately, some transmission lines are in greater demand than others and have become congested.[34] Even when some of a utility's transmission lines are in a critical "bottleneck" area, it cannot charge a higher price for the use of these congested lines than for the use of other lines the utility owns.[35] This restriction destroys any incentive for utilities to invest in critically needed new transmission capacity.[36]

Monopoly utilities once had a stable regulatory environment, a proven business structure and were therefore relatively risk-free investments, but this is no longer the case.[37] Transmission access rates set by the FERC do not include an appropriate risk premium for the utility companies.[38] The regulatory system is currently in a state of flux and the long-term profitability of transmission companies is uncertain.[39] Because the pricing mechanism does not account for this added risk, transmission owners are systematically undercompensated and are deterred from making further investments.[40] This problem is worsened because the FERC prices transmission on a short-run marginal cost basis that does not account for the massive start-up costs inherent in electricity transmission.[41] As a result, utilities cannot recover the

---

29. *Id.*

30. *See id.* at 493.

31. For an argument that regulatory failure on the part of the FERC is the main cause of the transmission problems of the last two decades, see Lawrence J. Spiwak, *You Say ISO, I Say Transco, Let's Call the Whole Thing Off*, PUB. UTIL. FORT., Mar. 15, 1999, at 38.

32. Promoting Wholesale Competition Through Open Access Non-Discriminatory Transmission Services by Public Utilities; Recovery of Stranded Costs by Public Utilities and Transmitting Utilities, 61 Fed. Reg. 21,540 (May 10, 1996) (issued Apr. 24, 1996) (codified throughout 18 C.F.R. pts. 35, 385 (2005)) [hereinafter Order 888].

33. Mateer, *supra* note 14, at 813. This rate is known as a "pro forma [transmission] tariff." 18 C.F.R. § 35.28(c)(1)(i) (2005); Spiwak, *supra* note 31, at 41.

34. AMY ABEL, ELECTRIC RELIABILITY: OPTIONS FOR ELECTRIC TRANSMISSION INFRASTRUCTURE IMPROVEMENTS 4-5 (2005), Cong. Res. Serv. (CRS) No. RL32075, *available at* http://www.opencrs.com/rpts/RL32075_20050610.pdf [hereinafter ELECTRIC RELIABILITY].

35. Mateer, *supra* note 14, at 814 (citing 18 C.F.R. § 35.28(c)(1)(i) (2005); Spiwak, *supra* note 31, at 41).

36. *Id.*

37. *See* ELECTRIC RELIABILITY, *supra* note 34, at 8-11.

38. Mateer, *supra* note 14, at 814-15.

39. *Id.* at 814.

40. *Id.* at 814-15.

41. Spiwak, *supra* note 31, at 43.

full cost of transmission assets.[42]

Many utility companies still "own" both transmission and generation assets.[43] New transmission lines will make it easier for merchant generators to enter the market and compete with such utilities in electric power generation, thus lowering the value of the utility's existing generation assets. As a result, when open access is mandated to allow for competitive merchant power generation, a utility that owns both generation and transmission infrastructures will under-invest in new transmission capacity and engage in entry-deterring practices to protect its existing assets.[44] After the FERC issued Order 888,[45] which mandated open access to transmission lines, investment in new bulk transmission facilities dropped by nearly 50%.[46]

Faced with this shifting, uncertain, and ineffective regulatory environment, many experts in the electric utility industry believe that substantial changes must be made to the current system of state siting regulation to promote critical transmission investments.

Under the existing regime:

> [A]ny state or local agency has the power to veto any proposed expansion of transmission capacity, and there is a large and growing shortage of transmission capacity. . . . [T]his allocation of regulatory power is certain to continue to produce a growing shortage of capacity that will have devastating effects on the price and availability of electricity.[47]

Others have questioned the role of state siting regulations in the shortage of electric transmission capacity. "State-based transmission siting processes vary considerably across the U.S., and, for the most part, worthy projects are approved, and deficient projects are discouraged, improved, or rejected. Most transmission projects are intrastate and small in scale."[48] "[T]ransmission continues to be built, especially to reinforce local areas. A somewhat old (1986) survey by the National Governor's Association found that only 18 of 533 transmission projects applied for at the state level were denied."[49] If, as these experts suggest, the current system of state siting is working well, dramatic changes to the system may be counterproductive.

---

42. *Id.*

43. *See* Mateer, *supra* note 14, at 817 (stating that "utility companies use subtle mechanisms to favor their own in-house generation firms").

44. Spiwak, *supra* note 31, at 43.

45. Order 888, *supra* note 32.

46. Spiwak, *supra* note 31, at 39.

47. Pierce, *supra* note 10, at 493.

48. DAVID H. MEYER & RICHARD SEDANO, TRANSMISSION SITING AND PERMITTING E-26 (2001) (National Transmission Grid Study Issue Papers, Department of Energy, May 2002).

49. Peter Fox-Penner, *Easing Gridlock on the Grid: Electricity Planning and Siting Compacts*, ELECTRICITY J., Nov. 2001, at 11, 12.

B. The Energy Policy Act of 2005

The Energy Policy Act of 2005[50] passed both houses of Congress with substantial majorities,[51] and was signed into law by President Bush on August 8, 2005.[52] The Act creates new energy efficiency standards, gives incentives for energy conservation and renewable energy sources, and provides billions of dollars for the research and development of cleaner energy technologies. The new legislation also establishes an independent organization to improve the reliability of electricity and expands FERC authority to ensure open access to transmission lines and prevent market manipulation.[53] In addition to those widely publicized provisions, the Act reforms the siting process for electric transmission facilities, including backstop siting authority for the FERC,[54] and strongly supports the formation of Regional Transmission Organizations (RTOs).[55]

At legislative hearings for the bill, many groups testified both in support of and in opposition to the proposed regulatory changes. Those supporting the changes included industry groups, such as the Edison Electric Institute (EEI) and American Public Power Association (APPA), and consumer groups, such as the Electric Consumers Alliance (ECA).[56] These organizations contend that the reform of siting procedures is necessary to ensure reliable, low-cost electric service for households and businesses, and to promote national security.[57] They assert that current siting procedures are stifling the growth of transmission facilities, particularly inter-state transmission lines, which are needed to relieve "bottle-neck" electric congestion areas and lower prices.[58]

Those groups opposing the bill's provisions include the Western Governors Association (WGA), the National Conference of State Legislatures,

---

50. 2005 Act, *supra* note 1.

51. 151 Cong. Rec. S9374 (2005) (stating the Senate agreed to a conference report on July 29, 2005 by recorded vote: 74-26 (Rollcall Vote No. 213)); 151 Cong. Rec. H6972 (2005) (stating the House agreed to a conference report on July 28, 2005 by recorded vote: 275-156 (Roll No. 445)).

52. 2005 Act, *supra* note 1.

53. *See* discussion *infra* Part III.B.

54. 2005 Act, *supra* note 1, § 216(b)(1); *see* discussion *infra*, Part III.C.2; *see also* Thomas R. Kuhn, *Who's Minding the Grid?* PUB. UTIL. FORT., Jan. 1, 2002, at 10, 10 (using term "backstop siting authority").

55. *See* discussion *infra* Parts III.C.4; III.C.2.

56. *See Electricity Proposals and Electric Transmission and Reliability Enhancement Act of 2003: Hearing Before the S. Comm. on Energy & Natural Res.*, 108th Cong. 74-80, 97-99 (2003) [hereinafter *Hearings*] (statements of H. Allen Franklin, Chairman, President, and CEO, Southern Company, on Behalf of Edison Electric Institute (EEI), and Alan H. Richardson, President and CEO, American Public Power Association (APPA)); MEYER & SEDANO, *supra* note 48, at E-28.

57. *See Hearings*, *supra* note 56, at 74-80, 98-99; MEYER & SEDANO, *supra* note 48, at E-28.

58. *Id.*

the National Association of Regulatory Utility Commissioners (NARUC), the National Association of Towns and Townships, the National Association of State Utility Consumer Advocates, and other state and local organizations.[59] These groups hope to retain their current siting power and argue that no changes are needed because the current methods for siting approval are being successfully adapted to the new challenges posed by regional markets.[60] They dispute the need to shift siting authority to regional or national entities, and argue that such a change might be counterproductive.[61]

In light of these conflicting views, an analysis of the current legal structure for the siting of electric transmission facilities is necessary. The need for, and potential impact of, proposed regulatory changes must be considered in light of the benefits and detriments associated with current procedures.

C. Critical Infrastructure

Considering the potential economic ramifications, an overloaded transmission grid may present an appealing target for terrorists.[62] "We are dependent on electricity for almost every aspect of daily life. The transmission grid, due to its interconnected properties and overstressed condition, is vulnerable to an intentional attack. . . . [W]idespread loss of grid functions would . . . be devastating."[63]

In 2001, "[t]he electric industry service[d] almost 130 million households [and] institutions and the United States consumed nearly 3.6 trillion kilowatt hours."[64] Electricity is necessary not only for most productive activities, but also for the production of other forms of energy such as oil.[65] "Were a

---

59. *See Hearings*, supra note 56, at 13-16, 24, 26-27, 216-21 (statements of David A. Svanda, President, National Association of Regulatory Utility Commissioners (NARUC), Gerald Norlander, Chairman, Electricity Committee, National Association of State Utility Consumer Advocates (NASUCA), and Bill Richardson, Governor of New Mexico, on behalf of the Western Governors' Association (WGA)); MEYER & SEDANO, *supra* note 48, at E-16, E-28.

60. *Id.*

61. *Id.*

62. Joel B. Eisen, *Regulatory Linearity, Commerce Clause Brinkmanship, and Retrenchment in Electric Utility Deregulation*, 40 WAKE FOREST L REV. 545, 557 n.63 (2005). *See also* ICF CONSULTING, THE ECONOMIC COST OF THE BLACKOUT: AN ISSUE PAPER ON THE NORTHEASTERN BLACKOUT, AUGUST 14, 2003 (2003), *available at* http://www.icfcon .com/Markets/Energy/doc_Files/blackout-economic-costs.pdf (stating that "[a] coordinated attack on a power grid could lead to . . . significant economic damages, both as a direct consequence of the attack, as well as through the ripple effects resulting from the strong inter-linkages between industry sectors").

63. Eisen, *supra* note 62, at 557 n.63.

64. THE NATIONAL STRATEGY FOR THE PHYSICAL PROTECTION OF CRITICAL INFRASTRUCTURES AND KEY ASSETS 50 (2003), *available at* http://www.dhs.gov/interweb/ assetlibrary/Physical_Strategy.pdf [hereinafter CRITICAL INFRASTRUCTURES].

65. *Id.*

widespread or long-term disruption of the power grid to occur, many of the activities critical to our economy and national defense—including those associated with response and recovery—would be impossible."[66]

In the aftermath of the August 2003 blackout, which affected the eastern part of the United States and Canada, there was speculation that the grid failure was caused or exacerbated by terrorism.[67] Although this was not the case,[68] investigation following the incident revealed a troubling vulnerability to future attacks.[69]

North America's energy sector relies increasingly on information technology and computer software to manage the complex electric transmission grid.[70] This reliance renders the grid vulnerable to "malicious cyber event[s]," defined by the Blackout Report "as the manipulation of data, software or hardware for the purpose of deliberately disrupting the systems that control and support the generation and delivery of electric power."[71] Some computer systems, designed in isolation like the local transmission grids they controlled, were developed without concern for cyber security and are now "directly . . . connected to the global Internet."[72]

Coinciding with this increased vulnerability to attack is an elevated threat level. "[T]he threat environment is changing and . . . the risks are greater than in the past."[73] "Current assessments suggest that there are terrorists and other malicious actors who have the capability to conduct a malicious cyber attack with potential to disrupt the energy infrastructure."[74] "The generation and delivery of electricity has been, and continues to be, a target of malicious groups and individuals intent on disrupting this system."[75]

At the same time, the deregulation of the electricity market has induced a variety of new firms to enter the industry.[76] In contrast to the large, strictly regulated monopolies of old, these new participants are varied in size, organization, and focus.[77] They purchase the level of security they can afford that is consistent with their organizational philosophy.[78] In addition, these varied stakeholders often fail to efficiently share the data needed to properly

66. *Id.*

67. *See* U.S.-CANADA POWER SYSTEM OUTAGE TASK FORCE, FINAL REPORT ON THE AUGUST 14, 2003 BLACKOUT IN THE UNITED STATES AND CANADA: CAUSES AND RECOMMENDATIONS 131 (2004), *available at* https://reports.energy.gov/ [hereinafter 2003 BLACKOUT REPORT].

68. *Id.* at 131, 135-36.

69. *Id.* at 132-33, 139-40.

70. *Id.* at 132.

71. *Id.*

72. *Id.* at 133.

73. *Id.*

74. *Id.* at 135.

75. *Id.* at 132.

76. CRITICAL INFRASTRUCTURES, *supra* note 64, at 51.

77. *Id.*

78. *Id.*

analyze and safeguard the systems they control.[79] Yet these firms may be responsible for key national assets and critical infrastructure.

Even worse, the potential damage of an attack on the nation's electricity infrastructure is magnified by the overloaded condition of the physical facilities that make up the national electricity grid.

> Over the past decade or more, electricity demand has increased and the North American interconnections have become more densely woven and heavily loaded, over more hours of the day and year. In many geographic areas, the number of single or multiple contingencies that could create serious problems has increased. Operating the grids at higher loadings means greater stress on equipment and a smaller range of options and a shorter period of time for dealing with unexpected problems.[80]

In order to alleviate this dangerous stress, new electric transmission lines must be sited and constructed.[81] However, companies seeking to expand transmission capacity face several difficulties. Transmission lines are long-term investments offering uncertain rates of return, and are easily derailed by either NIMBY opposition, state siting decisions, or the use of eminent domain.[82]

II. PRESENT LAW IS INADEQUATE TO DEAL WITH SITING PROBLEMS

A. Siting Law and Doctrine are Based on the Former Local Utility Monopoly Model

Electricity markets in the United States were once primarily local and isolated from one another. Increasingly, however, utility markets and companies are evolving into interdependent regional entities whose performance has a profound impact on the national economy.[83] At one time, utility companies held regulated monopoly status in their designated service areas.[84] They were responsible for both the generation and distribution of power within that area.[85] That situation, however, has undergone a fundamental change:

> In the United States, wholesale power sales by independent power producers (IPPs) began after passage of the Public Utility Regulatory Policy Act of

---

79. *Id.*

80. 2003 BLACKOUT REPORT, *supra* note 67, at 139-40.

81. CRITICAL INFRASTRUCTURES, *supra* note 64, at 51.

82. *Id.*

83. *See* MEYER & SEDANO, *supra* note 48, at E-14; Ashley C. Brown & Damon Daniels, *Vision Without Site: Site Without Vision*, ELECTRICITY J., Oct. 2003, at 23, 23.

84. Hoang Dang, *New Power, Few New Lines: A Need For A Federal Solution*, 17 J. LAND USE & ENVTL. L. 327, 330 (2002).

85. *Id.*

> 1978, which established the right of non-utility producers to operate and sell their energy to utilities. This led to extensive IPP development in the northeast and west, increasing in-region and inter-regional power sales as utility loads grew without corresponding utility investments in transmission.[86]

The share of power that investor-owned utilities purchased from other utilities and IPPs has increased from 17.8% in 1989 to 37.3% in 2002, and the share of power purchased by large public power entities increased from 36.3% in 1992 to 40.5% in 2002.[87]

In the Energy Policy Act of 1992,[88] Congress enhanced competition by introducing exempt wholesale generators that would compete with traditional utility generators in wholesale electric markets.[89] The Act also broadened the FERC's authority to order access to transmission lines on a case-by-case basis.[90] Further, in Order 888,[91] the FERC required open access of public utility-owned, operated, or controlled interstate transmission facilities for energy sales.[92] The Order also allowed utilities to fully recover "stranded" costs; that is, costs that the utility "prudently incurred" to provide service to customers under the old regime of regulatory rate setting, and which are no longer recoverable because market-based open access has allowed those customers to obtain their service from a competitor.[93] In turn, an increasing number of companies became involved in the wholesale production and sale of electric power to the areas that needed it most.[94]

"Merchant generators" of electric power, those built solely for market sales and not to serve the needs of a specific locale, have proliferated. At the same time, however, transmission capacity has been stagnant, increasing at a rate of less than 1% per year.[95] The result is an antiquated transmission system overloaded by the transfer of newly generated wholesale power.[96] Several states have imposed temporary moratoria on the construction of new merchant plants in part because of concern that the local transmission grids will be overwhelmed by any influx of new power.[97] "[T]he increased loads and flows across a transmission grid that has experienced little new investment is causing

---

86. 2003 BLACKOUT REPORT, *supra* note 67, at 32.

87. *Id.* (citing RDI PowerDat database).

88. Energy Policy Act of 1992, Pub. L. No. 102-486, 106 Stat. 2776 (codified as amended in scattered sections of U.S.C.) (1992).

89. 15 U.S.C. § 79z-5a (2000).

90. 16 U.S.C. § 824j(a) (2000).

91. Order 888, *supra* note 32.

92. Eisen, *supra* note 62, at 550.

93. DAVID J. MUCHOW & WILLIAM A. MOGEL, ENERGY LAW & TRANSACTIONS § 82.03(4) (2004).

94. Dang, *supra* note 84, at 327-28.

95. ELECTRIC RELIABILITY, *supra* note 34, at 5.

96. Dang, *supra* note 84, at 328, 332; Eisen, *supra* note 62, at 555-56.

97. *Nervous of NOx, Southern Govs. Put Plants on Hold*, 17 ELECTRICITY DAILY 40 (Aug. 28, 2001).

greater 'stress upon the hardware, software and human beings that are the critical components of the system.'"[98]

This strain on the transmission grid has already had significant economic consequences, leading to blackouts and power shortages in the bottleneck areas most lacking transmission capacity.[99] Because electricity moves along all available paths once introduced into the power grid, every flow of electricity affects the entire distribution network.[100] The transfer capacity of an overall transmission system is therefore limited to the capacity of the part of the system where electricity is most narrowly channeled.[101] As a result, problems in bottleneck areas can have wide-ranging effects. In several instances, large areas lost power because of grid problems.[102] One notable event, the blackout of August 2003 that extended across much of the northeastern United States and Canada, affected over 50 million people and caused an economic loss of between $7 and $10 billion.[103] A bi-national task force was appointed to ascertain the causes of the blackout and a means of avoiding its repetition.[104] The task force's comprehensive report noted a decline in the United States' relative electric transmission capacity:

> Competition is not the only thing that has grown over the past few decades. Between 1986 and 2002, peak demand across the United States grew by 26%, and U.S. electric generating capacity grew by 22%, but U.S. transmission capacity grew little beyond the interconnection of new power plants. Specifically, "the amount of transmission capacity per unit of consumer demand declined during the past two decades and . . . is expected to drop further in the next decade."[105]

Unfortunately, all indications are that construction of transmission capacity will continue to lag behind surging generation and transmission demands.[106]

Many experts believe that the present procedure for siting transmission facility projects slows their construction.[107] Although transmission technology and the character of the markets have changed, the method of obtaining approval for siting new transmission projects has remained largely in a "time

---

98. 2003 BLACKOUT REPORT, *supra* note 67, at 32 (quoting a letter from Michael H. Dworkin, Chairman, Vermont Public Service Board, Feb. 11, 2004, to Alison Silverstein and Jimmy Glotfelty).

99. Dang, *supra* note 84, at 332-33.

100. Sager A. Williams, Jr., Comment, *Limiting Local Zoning Regulation of Electric Utilities: A Balanced Approach in the Public Interest*, 23 U. BALT. L. REV. 565, 572 (1994).

101. *Id.*

102. *See* 2003 BLACKOUT REPORT, *supra* note 67, at 103.

103. ICF CONSULTING, *supra* note 62.

104. *See* 2003 BLACKOUT REPORT, *supra* note 67, at 131.

105. *Id.* at 32 (quoting HIRST, *supra* note 8, at vii).

106. Eisen, *supra* note 62, at 556.

107. *See, e.g.*, Dang, *supra* note 84, at 329, 339; Fox-Penner, *supra* note 49, at 13.

warp."[108] States have exclusive jurisdiction over transmission facility siting and each state has its own method of reviewing proposed transmission projects to determine whether a project is necessary and preferable to alternative designs.[109] Any federal agencies impacted because of environmental concerns or the desired use of federal lands must conduct their own reviews.[110]

For an interstate transmission project, each affected state must approve the project.[111] Before construction can begin, a large interstate project may have to successfully navigate several independent state and federal review procedures, a process that has taken up to a decade and caused considerable expense.[112] New transmission projects, particularly interstate ones, are already complex and expensive long-term investments.[113] Under the current siting regime, such projects are further complicated during the approval process by local politics, by interstate squabbles, or by state courts empowered to consider only intrastate concerns in granting siting approval or the use of eminent domain.

B. Showing of Need and Public Use for Eminent Domain Purposes

New transmission lines must generally run across lands owned by private individuals or firms, or by state or federal agencies. Therefore, electric utilities must obtain easements in order to begin construction.[114] Because power lines must be aligned across the lands of multiple owners, obtaining easements through individual negotiations would render the electric companies vulnerable to holdout problems.[115] Eminent domain has traditionally been

---

108. Brown & Daniels, *supra* note 83, at 24.

> [S]iting and eminent domain[] has undergone dramatically little change in the past decade. While policy has promoted competition in regional bulk power markets and removal of entry barriers, the siting laws and eminent domain statutes have continued for the most part in a time warp, unchanged from the days of local monopolies.

Id.

109. Dang, *supra* note 84, at 335-36; *infra* discussion Part II.C.

110. *See, e.g.*, Classes of Actions that Normally Require EISs, 10 C.F.R. pt. 1021(D) app. D, D5, D6 (2005) (requiring environmental impact statements for transmission system additions and integration); Permits for Structures or Work In or Affecting Navigable Waters of the United States, 33 C.F.R. § 322.5(h)(5)(i) (2005) (requiring a federal permit for transmission lines crossing navigable waters); *see also* Fox-Penner, *supra* note 49, at 13 (stating that if Federal waters or lands are involved in the siting, "a number of federal approvals may be needed from a disparate set of agencies").

111. *See* MEYER & SEDANO, *supra* note 48, at E-5, E-15; Dang, *supra* note 85, at 339; Fox-Penner, *supra* note 49, at 13.

112. *See* MEYER & SEDANO, *supra* note 48, at E-8 to 9, E-38 to 39.

113. MEYER & SEDANO, *supra* note 48, at E-42 (asserting that "[d]elay and controversy have been more common in larger, interstate projects"); Eisen, *supra* note 62, at 556-57.

114. *See* Lisa M. Bogardus, *Recovery and Allocation of Electromagnetic Field Mitigation Costs in Electric Utility Rates*, 62 FORDHAM L. REV. 1705, 1707-08 (1994).

115. *See, e.g.*, Richard A. Epstein, *Rights and "Rights Talk"*, 105 HARV. L. REV. 1106,

used to acquire necessary easements; the power of eminent domain, however, can only be exercised for public use.[116] Therefore, in order to obtain approval for a project, there must be a showing of public need. In the past, the public need requirement was satisfied by a showing that the project would increase the reliability of the local transmission grid, to the benefit of the local consumers. Given the change in the electricity generation and transmission market, the benefit of many transmission facilities now inures primarily to a regional energy market instead of a locality. For example:

> Due to the dynamic and highly integrated nature of the AC grid, an upgrade in one state may be required to enhance reliability and relieve congestion in an adjacent state. Also, a transmission addition may be required in a state to enable an upgrade undertaken in an adjoining state to function as planned.[117]

While the need for siting transmission lines is regional and national, courts generally act on the proposition that a State cannot use its power of eminent domain for the benefit of the citizens of another State.[118] Courts find this limitation within the source of the legislative power; the sovereign is obligated to protect and promote the health, safety, morals, and welfare of citizens of the individual state.[119]

1114 (1992) (reviewing MARY ANN GLENDON, RIGHTS TALK: THE IMPOVERISHMENT OF POLITICAL DISCOURSE (1991)) (asserting that "property rights are not absolute, for individual owners are not allowed to hold out against the community at large if the community is prepared to pay them just compensation for their losses").

116. *See* Kelo v. City of New London, 126 S. Ct. 2655 (2005) (adopting broad view of "public use" as including "public purpose"). *Kelo* affirmed that a locality "would no doubt be forbidden from taking petitioners' land for the purpose of conferring a private benefit on a particular private party." *Id.* at 2661.

117. TRANSMISSION ACCESS POLICY STUDY GROUP, EFFECTIVE SOLUTIONS FOR GETTING NEEDED TRANSMISSION BUILT AT REASONABLE COST 6 (2004) [hereinafter TAPS], *at* http://www.tapsgroup.org/sitebuildercontent/sitebuilderfiles/effectivesolutions.pdf.

118. *See* Adams v. Greenwich Water Co., 83 A.2d 177, 182 (Conn. 1951) (holding that "no state is permitted to exercise or authorize the exercise of the power of eminent domain except for a public use within its own borders"); Clark v. Gulf Power Co., 198 So. 2d 368, 371 (Fla. Dist. Ct. App. 1967) (holding that the power of eminent domain is only "for the use and benefit of the people within the state"); People *ex rel.* Trombley v. Humphrey, 23 Mich. 471, 475 (Mich. 1871) (holding that "[f]or the one to enter the sphere of the other and employ its officers and machinery in the exercise of its eminent domain for the benefit of the other would not only be as much without warrant, but also as much a work of supererogation, as for the United States to exercise the like authority and employ the like agencies in a foreign country"); Miss. Power & Light Co. v. Conerly, 460 So. 2d 107, 112-13 (Miss. 1984) (excluding benefits to be derived outside the state from consideration in eminent domain proceedings); Grover Irrigation & Land Co. v. Lovella Ditch, Reservoir & Irrigation Co., 131 P. 43, 55 (Wyo. 1913) (stating that "in every case where the use as a justification for the proceeding has been questioned, the inquiry in that respect has been confined to the interest and welfare of the state or sovereignty within whose limits or jurisdiction the land sought to be condemned is located").

119. *See, e.g.*, Kohl v. United States, 91 U.S. 367, 373-74 (1875) (holding that "the right

In *Mississippi Power & Light Co. v. Conerly*,[120] for example, the Mississippi Supreme Court dismissed out-of-state benefits as a possible basis for the use of eminent domain and held that Mississippi citizens must be the primary beneficiaries of any such use.[121] The court interpreted the statutory language of "public necessity" and "public use" to refer to use by the citizens of the state of Mississippi.[122] The Mississippi legislature later endorsed the court's interpretation when, in 1992, it amended title 77, chapter 3, section 14 of the Mississippi Code to specifically consider only benefits to Mississippi citizens in deciding whether to construct new transmission facilities.[123] From the determination that eminent domain cannot be used to benefit citizens of another state flows the question of whether eminent domain can ever be used for projects that somehow benefit citizens of another state as well as the citizens of the state exercising the power.

When a use indirectly and necessarily benefits citizens of another state as a byproduct of its use by citizens of the state that is exercising the power of eminent domain, courts have universally held that such a collateral benefit does not preclude that state from proceeding in its condemnation.[124] A trickier question arises, however, with eminent domain projects that are intended to benefit the citizens of two or more states, but could be scaled down to meet the needs of only the state exercising the power. If the proposed project could be conceived as a combination of two projects, one of which benefits only the citizens of the state in which it is to be constructed, while the other benefits

---

of eminent domain . . . is a right belonging to a sovereignty to take private property for its own public uses, and not for those of another"); Square Butte Elec. Coop. v. Hilken, 244 N.W.2d 519, 525 (N.D. 1976) (emphasizing that "the public benefit, while not confined exclusively to the state authorizing the use of the power . . . is nonetheless inextricably attached to the territorial limits of the state because the state's sovereignty is also so constrained").

120. 460 So. 2d 107 (Miss. 1984).

121. *Id.* at 113.

122. *Id.* (referring specifically to MISS. CODE ANN. § 11-27-15).

123. MISS. CODE ANN. § 77-3-14 (1972) ("The commission shall develop . . . an analysis of the . . . needs for expansion of facilities for the generation of electricity in Mississippi . . . to achieve maximum efficiencies for the benefit of the people of Mississippi").

124. *See* Columbus Waterworks Co. v. Long, 25 So. 702, 703 (Ala. 1899) (holding that the right of condemnation "is not to be denied where public uses are" promoted in other states as well as the one granting the right); Gilmer v. Lime Point, 18 Cal. 229, 253 (Cal. 1861) (holding that benefits to citizens of Oregon do not make an improvement "less [of] a public use in California"); *Adams*, 83 A.2d at 182 (stating that "[i]f the taking is for a public use which will provide a substantial and direct benefit to the people of the state which authorizes it, it is a proper exercise of the power of eminent domain even though it also benefits the residents of another state"); Wash. Water Power Co. v. Waters, 115 P. 682, 686 (Idaho 1911) (holding that incidental benefits to citizens of other "state[s] will not defeat the right of condemnation"); *see also* Gralapp v. Miss. Power Co., 194 So. 2d 527, 531 (Ala. 1967) (stating that "the right to condemn in this case cannot be denied because public uses in another state would be promoted also"); *Clark*, 198 So. 2d at 371 (holding that land cannot be condemned solely for the benefit of another state); *Square Butte Elec. Coop*, 244 N.W.2d at 525 (holding that "other states may also be benefited" [sic]).

only individuals outside the state, it makes little sense to allow the power to be used for the merged project. Determining what portion of the project benefits which citizens is not easy, however, and courts have held that the use of eminent domain for such projects is constitutional.[125]

Although it is clear from recent cases that courts recognize the regional benefits of some interstate transmission proposals and want to consider these benefits when determining whether to use eminent domain,[126] courts are tied to the language of intrastate benefits. In order to approve an interstate project, the courts must find that it will result in at least substantial intrastate benefits. Recently, however, courts have been increasingly confronted with proposals for facilities that are aimed at improving the regional energy market, under which few direct benefits accrue within the state. Even if the proposed construction is in a critical regional electricity bottleneck and is needed for reasons of national security, the court must find that substantial intrastate benefits arise from the proposal in order to approve the use of the state's power of eminent domain.

Courts were first presented with this problem in the 1960s, when they began to issue decisions on the extent to which eminent domain could be used to acquire rights-of-way for transmission facility projects having significant regional impacts. Courts have placed varying requirements upon such uses of power, from requiring that the project's primary beneficiaries be in-state citizens, to merely requiring the accrual of a substantial benefit to in-state citizens.[127] These cases seem to focus on whether it is constitutional to use the power of eminent domain for purposes that benefit citizens of other states to varying degrees. Rather than being a factor supporting the use of eminent domain, benefits to other states are generally a barrier to a finding of public use. Some states, however, have departed from *Conerly*, choosing to view benefits to regional energy markets as positive factors in determining whether a project constitutes a public use.

Some courts interpret in-state benefits broadly, such as by including

125. *See, e.g.*, *Adams*, 83 A.2d at 182 (rejecting the argument that the right of eminent domain should be limited to the portion of a project that is required for the citizens within the state).

126. *See* Grand Canyon Trust v. Ariz. Corp. Comm'n, 107 P.3d 356, 364 (Ariz. Ct. App. 2005) (holding that utility customers included wholesale customers whose retail users were not necessarily in Arizona); Cross-Sound Cable Co. v. Rocque, No. CV030473582, 2003 WL 1900775, at *18 (Conn. Super. Ct. Apr. 9, 2003) (balancing regional needs and environmental issues); Neb. Pub. Power Dist. v. Johnson, No. A-97-114, 1998 WL 765718, at *6 (Neb. Ct. App. Sept. 22, 1998) (noting the utility's "obligations as a member of the Mid-Continent Area Power Pool").

127. *See Gralapp*, 194 So. 2d at 531 (stating that "private property may not be condemned unless it is to be subjected to a recognized public use, affording benefits which are not vague, indefinite or restrictive"); *Adams*, 83 A.2d at 182 (requiring "substantial and direct benefit[s]"); *Clark*, 198 So. 2d at 371 (holding that the required public use must be well-defined and "within the control of the state"); *Grover Irrigation & Land Co.*, 131 P. at 56 (finding indirect benefits insufficient).

"public advantage" within the concept of "public use." In *Square Butte Electric Cooperative v. Hilken*,[128] a group of rural electricity cooperatives entered into an agreement with a Minnesota electric company to provide it with electric power in return for financing the construction of new facilities.[129] The group sought the use of eminent domain to construct a transmission line from the facilities to the Minnesota company's plant.[130] The affected landowners challenged this taking as lacking a valid public use. The trial court refused to take into account the advantage that increased reserve and emergency power supplies would confer upon the state. The Supreme Court of North Dakota reversed, holding that the proper test was whether the alleged benefits of the proposed facility would "provide . . . either singly or in unison, a substantial and direct benefit to North Dakota."[131] The court decided that, although insufficient as individual factors, increased reserve and emergency power supplies, the stabilizing effect of the proposed line on the existing system, the possibility that the project would provide electricity to North Dakota consumers in the future, and reduced future power costs satisfied that test in combination.[132] A strong concurrence stated that while benefits to North Dakotans were clearly indirect, indirect benefits were enough to constitute a constitutionally valid public use.[133]

Several other courts have issued similar pragmatic decisions taking regional benefits into account.[134] This progressive regional approach, with its lower standard for generating easements for electric transmission lines, appears to be more prevalent in the Western states.[135] Perhaps in part because of this Western legal tradition, the Western Governor's Association has been a leading opponent of the proposed changes to siting authority. It insists that Western states have a long history of cooperating and approving necessary interstate projects reasonably quickly.[136] The Wyoming Public Service Commission has even suggested that RTOs are unnecessary because the

128. 244 N.W.2d 519 (N.D. 1976).

129. *Id.* at 521-22.

130. *Id.*

131. *Id.* at 525.

132. *Id.* at 525-31.

133. *Id.* at 532-33 (Pederson, J., concurring).

134. *See, e.g.*, Oxendine v. Pub. Serv. Co. of Ind., 423 N.E.2d 612, 615-17 (Ind. Ct. App. 1980) (interpreting statutory language to include out-of-state wholesale customers in the definition of "the public"); Mont. Power Co. v. Bokma, 457 P.2d 769, 772-73 (Mont. 1969) (finding the public's right to use power, if necessary, to constitute a public advantage sufficient for the exercise of eminent domain).

135. *See, e.g., Mont. Power Co.*, 457 P.2d at 772 (stating that "the broad view, essentially requires only a use conferring a 'public advantage' or a 'public benefit.' Montana, as with many western states, has adhered to the broad view since 1895, presumably to promote general economic development").

136. *Hearings*, *supra* note 56, at 216, 220-21 (statement of Bill Richardson, Governor of New Mexico, on behalf of the Western Governors' Association).

Western states already constitute a single unified wholesale market.[137] Because these states have large unpopulated land areas and contain a significant amount of federal land, they have indeed been forced to cooperate to a greater extent than most Eastern states.[138] This may have played a role in the evolution of their eminent domain jurisprudence.

One jurisdiction has ruled that regional integration is a direct intrastate benefit. In *Stone v. Pennsylvania Public Utility Commission*,[139] the Superior Court of Pennsylvania held that it was proper to use the state's power of eminent domain for the construction of a high-power transmission line connecting the electric grids of one of its principal cities to that of another state.[140] The court stated that "[o]ne of the principal considerations of public convenience and necessity is the need for integration of the bulk power transmission systems of Philadelphia and Baltimore."[141] The court relied on benefits from greater economies of scale and improved grid reliability in holding that electric grid integration was a major benefit for Pennsylvania citizens.[142] The court also noted the public need for alternate sources of electric power in case of a national emergency.[143] The holding in *Stone* has since been followed and reinforced in Pennsylvania.[144]

Finally, one court adopted a position in direct opposition to *Conerly*. In *Gralapp v. Mississippi Power Co.*,[145] the Supreme Court of Alabama ruled that it was appropriate to use the power of eminent domain for the construction of a transmission line connecting the Alabama and Mississippi electric grids, where nearly all of the power sent over the line flowed out of Alabama.[146] The company building and operating the line was a Mississippi corporation that was not "subject to the jurisdiction of the Alabama Public Service Commission," and the only power flowing into Alabama was the result "of a contractual arrangement."[147] Thus, the State of Alabama had no right to control the line and no guarantee that the line would ever directly benefit Alabama's citizens.[148] Nevertheless, the court found a public use, essentially holding that *any* benefit to the citizens of the State of Alabama was sufficient to support the use of eminent domain.[149] This result diametrically opposes the

137. Bruce W. Radford, *Electric Transmission: Do State Regulators Still Have a Voice?*, PUB. UTIL. FORT., Nov. 15, 1999, at 42, 44.

138. *See* MEYER & SEDANO, *supra* note 48, at E-19, E-37.

139. 162 A.2d 18 (Pa. Super. Ct. 1960).

140. *Id.* at 19-22.

141. *Id.* at 21.

142. *Id.*

143. *Id.*

144. *See, e.g.*, Dunk v. Pa. Pub. Util. Comm'n, 232 A.2d 231, 234-36 (Pa. Super. Ct. 1967).

145. 194 So. 2d 527 (Ala. 1967).

146. *Id.* at 530-31.

147. *Id.* at 530.

148. *Id.*

149. *Id.* at 531.

rule from *Conerly* that the primary beneficiaries of a project must be in-state citizens in order for a state to exercise its power of eminent domain. Under *Gralapp*, the use of eminent domain could be justified for nearly any project through the simple addition of a contractual arrangement with a state utility company entitling that state to a small amount of power. Other states have vehemently refused to find a public use when small shares of power are allocated to in-state use.[150]

Many courts have not made the leap to approving the use of eminent domain based on a finding that purely interstate transmission lines give rise to sufficient public benefits.[151] The extent of a possible trend towards a more liberal interpretation of public use and a more liberal use of eminent domain for the siting of critical electric transmission facilities remains unclear. There is no chronological progression from *Conerly* to *Gralapp*. Rather, many of the most definitive cases were decided more than 20 years ago. The importance of regional energy markets has grown dramatically since those decisions, yet they are not generally being reevaluated. Eminent domain questions, however, are mainly issues of constitutional interpretation or fundamental legal principles, and are not influenced by legislative or policy changes. Eminent domain jurisprudence may not have the ability to adapt to changing circumstances, and it seems unlikely that this area of the law will change dramatically any time soon. This adaptive failure is an important consideration for an industry desperately in need of large interstate transmission projects that surely will require the use of eminent domain.

C. Showing of Need and Public Use for Permitting Purposes

In addition to obtaining government authority for the condemnation of necessary rights of way, utility companies must obtain state approval for the construction of new transmission facilities or the expansion of those already existing.[152] This permitting process is generally delegated to a state utility commission and is conducted in accordance with statutory standards and procedures.[153]

Typically, state commissions control entry into the public utility market through the grant of "certificates of public convenience and necessity."[154] These certificates are revocable licenses intended to avoid "unnecessary

150. *See, e.g.*, Tampa Elec. Co. v. Garcia, 767 So. 2d 428, 430, 436 (Fla. 2000) (finding that a contractual arrangement to provide 30 megawatts of a power plant's planned capacity of 514 megawatts to a local utility company was insufficient to show a state's need for the power plant).

151. *See Clark*, 198 So. 2d at 371 (stating that any benefits accruing to Florida citizens would be mere conjecture and could not support the taking of land for a transmission line); *Conerly*, 460 So. 2d at 112 (upholding lower court's determination that future benefits were "highly speculative); *see also* Brown & Daniels, *supra* note 83 (collecting other cases).

152. MUCHOW & MOGEL, *supra* note 93, at § 52.04(3).

153. *Id.*; *see also* Fox-Penner, *supra* note 49, at 13.

154. MUCHOW & MOGEL, *supra* note 93, at § 2.05(1).

duplication of . . . facilities" "by regulating competition between utilities within the same geographic area."[155] In order to gain approval, a utility must typically demonstrate that the planned expansion of transmission capacity satisfies a public need in that it is "'required by the present or future public convenience and necessity.'"[156] Whether there is a public necessity is determined by such factors as the proposal's feasibility and the estimated demand for electrical service.[157]

For interstate transmission projects, the determination-of-need component of the permitting process is similar to that in eminent domain proceedings. The main issue in both determinations is whether substantial benefits accrue within the state. At the least, a determination of public need requires that a transmission facility somehow benefits state citizens. Courts often find a need wherever they find a benefit, such as lower prices or improved reliability.[158]

In delineating the contours of the permissible exercise of eminent domain, legislatures are constrained by state and federal constitutional standards of "public use." In administering the permitting process, however, constitutional limitations on legislative authority are less stringent. The U.S. Constitution is one of enumerated powers; all powers not delegated to the national government are reserved to the states. Such reserved powers include the protection of the public health, safety, and welfare, and the enactment of social and economic legislation, which is given deferential review in federal courts. The United States Supreme Court has long ago affirmed the states' right to regulate private firms, such as utilities, that are "affected with a public interest."[159]

State regulation of the electric industry began early in the twentieth century.[160] At that time, power stations were located in cities and had to run distribution lines through "public thoroughfares" to reach their retail customers.[161] This interference, along with the enormous public demand for electricity, led state governments to conclude that the provision of electricity "was impressed with a public interest" and required regulation.[162] State legislatures began by regulating the industry directly, but soon found this approach inefficient.[163] State governments gradually formed regulatory

155. *Id.*

156. *Id.* at § 2.05(1) (quoting 18 C.F.R. § 157.6), 52.04(3); *see also* Fox-Penner, *supra* note 49, at 20-28; Bogardus, *supra* note 114, at 1706-07.

157. MUCHOW & MOGEL, *supra* note 93, at § 2.05(1).

158. *See, e.g.*, Sierra Pac. Power Co., 64 C.P.U.C.2d 442, 451-52 (1996), *available at* 1996 WL 37798; New England Elec. Transmission Corp., 48 P.U.R.4th 477, 483-84 (N.H.P.U.C. 1982).

159. Munn v. Ill., 94 U.S. 113, 126 (1877); *see also* MUCHOW & MOGEL, *supra* note 93, at § 2.03.

160. MUCHOW & MOGEL, *supra* note 93, at § 52.02.

161. *Id.*

162. *Id.*

163. *Id.*

commissions, which are the bodies responsible for utility regulation today.[164]

Regulation was originally intended to prevent an electric utility company's natural monopoly over its service area from resulting in "excessive profits."[165] "Over time . . . the scope of regulation has broadened to include consideration of environmental, financing, service, and economic subjects."[166] The public interest and general welfare, as prerequisites for permitting approval, are almost wholly in the legislative domain.

Although the public need determination in the permitting process is not arbitrary, it is flexible and responsive to changing political and economic conditions. Nevertheless, there are still many permitting processes that do not take regional benefits into account to the extent that the *Gralapp* court did in the eminent domain context forty years ago.[167] When making a determination of public need, there are some permitting processes that do not consider reasonable future expectations of need in the community or contractual agreements between the company proposing the new project and state utilities.[168] Yet these factors are often considered by courts in the context of eminent domain.[169]

For example, in *Point of Pines Beach Ass'n v. Energy Facilities Siting Board*,[170] the Supreme Judicial Court of Massachusetts overturned a decision to grant approval for a new power facility on the basis of a power purchase agreement.[171] The Board had decided that a contractual agreement with a local utility to purchase power generated by the new facility was prima facie evidence of need within the Commonwealth of Massachusetts.[172] The court disagreed, holding that there had to be actual evidence that the Commonwealth would need the power generated by the facility when it opened.[173] The court stated that the statistical forecasts provided could not prove whether the

---

164. *Id.*

165. *Id.*

166. *Id.*

167. *Compare Gralapp*, 194 So. 2d 527 (approving the use of eminent domain for a transmission line that sent nearly all of the power out of state), *with Tampa Elec. Co.*, 767 So. 2d 428 (rejecting the use of eminent domain to build a power plant that would provide only a small amount of power to a local utility company).

168. *See, e.g.*, Point of Pines Beach Ass'n v. Energy Facilities Siting Bd., 644 N.E.2d 221, 223-24 (Mass. 1995).

169. *See Gralapp*, 194 So. 2d at 530-31 (holding that a contractual arrangement for a local company to purchase power was sufficient to support a finding of public use); *Oxendine*, 423 N.E.2d at 616-17 (considering present and "reasonable future needs"); *Conerly*, 460 So. 2d at 112 (noting the lack of a contractual arrangement with an intrastate utility in rejecting the use of eminent domain); *Square Butte Elec. Coop.*, 244 N.W.2d at 527-30 (N.D. 1976) (considering projected energy needs in determining whether a public use exists and finding a public use partially on the basis of a contractual arrangement with a local utility).

170. 644 N.E.2d 221 (Mass. 1995).

171. *Id.* at 222, 224.

172. *Id.* at 222.

173. *Id.* at 223-24.

Commonwealth would need additional power and that the contract was not the product of market forces.[174] Therefore, it reversed the board's decision.[175] Although the case was deemed to be a matter of statutory interpretation, the actual language of the relevant statute required only that the new facility "provide a necessary energy supply for the commonwealth."[176] This type of interpretation, where courts decide that only specified intrastate needs meet the statutory requirements, is not uncommon.[177]

It is important to note, however, that two years later, the Massachusetts legislature enacted new legislation that removed the necessity requirement and replaced it with an explicit benefits analysis similar to that customarily found in eminent domain proceedings.[178] The law takes into account regional benefits insomuch as they contribute to the reliability of the energy supply of Massachusetts.[179] This legislative flexibility and evolution is impressive, and it suggests that state permitting processes can adapt to changing conditions.

Furthermore, approval processes generally seem more liberal than eminent domain determinations. In *Re Sierra Pacific Power Co.*[180], for example, the California Public Utilities Commission approved the construction of a new high-voltage transmission line, even though it provided only indirect benefits to California citizens.[181] Over ninety-two percent of the proposed transmission line's load would be attributable to the State of Nevada.[182] The Commission held that no weight could be given to the utility's increased import capacity resulting from the new transmission line, but that some weight could be given to the role of the project in providing new transmission capacity for citizens of other states.[183]

In addition, the court noted the importance of regional energy markets and its interest in mitigating the difficulties that the transmission grid currently faces.[184] Noting that several California blackouts resulted from problems in Nevada and from an overloaded grid, the Commission held that the possible facilitation of future interconnection and modest improvements in the reliability of service to California residents were sufficient to show public necessity for the project.[185] In finding public necessity, the commission considered many factors, including the following: "[1] community values, . . . [2] recreational and park areas, . . . [3] historical and aesthetic values, and .

174. *Id.* at 224.
175. *Id.* at 222-24.
176. *Id.* at 223-24 (quoting MASS. GEN. LAWS ANN. ch. 164, § 69J (West 2003)).
177. *See, e.g., Tampa Elec. Co.*, 767 So. 2d at 435-36.
178. MASS. GEN. LAWS ANN. ch. 164, § 69J 1/4 (West 2003).
179. *Id.*
180. 64 C.P.U.C. 2d 442 (1996), *available at* 1996 WL 37798.
181. *Id.* at 442.
182. *Id.* at 450, 472.
183. *Id.* at 452, 454, 476.
184. *Id.* at 469.
185. *Id.* at 451, 469.

. . [4] influence on the environment."[186] Because the Commission was given broad discretion in making its determination, it seems to have been able to take changing markets and corresponding needs into account in a way that courts considering the use of eminent domain power cannot. This is particularly apparent in the board's assertion that it considers benefits provided to citizens of other states.

The New Hampshire Public Utilities Commission takes an even broader view, equating regional benefits to intrastate ones in siting approval. In *Re New England Electric Transmission Corp.*,[187] a utility applied for a certificate of site and facility to construct a new transmission line that would create energy and cost savings in the entire New England region, mainly outside of New Hampshire.[188] The Public Utilities Commission noted that this was the first case it had seen that proposed a transmission line designed mainly to more efficiently convey electric power already in the regional grid, and not associated with any specific generator.[189] The main New Hampshire utility was a member of NEPOOL, a voluntary cooperative energy pool that ran many New England utilities in concert in order to achieve cost savings.[190] The Commission decided that, because of the interconnectedness between the New Hampshire and New England electric grids, benefits to New England were necessarily benefits to New Hampshire.[191] Evidence that the project would reduce the cost of electric power and improve the ability of utilities to deliver electric power in the New England area was sufficient to fulfill the statutory requirement that the project be "required to meet the present and future demand for electric power."[192] Several years later, the Commission conditioned approval of a similar project on evidence that the project would not jeopardize the reliability of other grid systems in the Northeast, even though it posed no danger to New Hampshire or New England.[193]

The fact that some states have equated, or almost equated, regional and state benefits, is very significant. As a result, those benefits are effectively internalized for the purposes of electric transmission permitting, which encourages an efficient level of new development. Although most jurisdictions have not completely equated regional and state benefits,[194] it is clear that a responsive legislature and court system can successfully adapt to the new reality of interconnected and interdependent regional transmission

186. *Id.* at 476.

187. 48 P.U.R.4th 477 (N.H.P.U.C. 1932).

188. *Id.* at 477, 478, 482.

189. *Id.* at 482.

190. *Id.* at 483.

191. *Id.* at 483-84.

192. *Id.* at 484, 491.

193. New England Hydro-Transmission Corp., 71 N.H.P.U.C. 727, 746 (1986) (interpreting statutory language that required a project not to adversely affect system stability).

194. *See, e.g.*, MEYER & SEDANO, *supra* note 48, at E-13.

grids.[195]

Despite the changes being made at the state level, the grid capacity crisis continues to develop and the problem is only getting worse.[196] If the current siting regime does not change, state courts' abilities to adequately account for regional benefits in the approval of future electric transmission facility projects will be critical.

D. Local Legislative Determinations

Many state and local government agencies, officials, and organizations have staunchly opposed any changes to the current system that would divest them of siting authority,[197] citing in part their unique ability to weigh the impact of transmission projects on local communities.[198] For siting decisions with purely local or intrastate impact, local officials may be best situated to balance local economic, environmental, and similar needs. In fact, local officials may be better able to measure local impact precisely because they are steeped in "local politics." To the extent that localities and individual states internalize the effects of their own siting decisions, they have an incentive to authorize an optimal pattern of new transmission infrastructure.

Many state siting boards, however, are not authorized to consider interstate benefits. One study indicates that in 22 states, localities can block interstate transmission expansion projects.[199] Also, local and state decisionmaking bodies naturally attempt to internalize the gains resulting from their decisions and externalize the corresponding costs. This leads to a lack of responsibility and concern for regional and national transmission needs. At present, there is widespread disagreement over whether this responsibility should belong to a local, state, or national entity.[200]

---

195. *See also* OHIO REV. CODE ANN. § 4928.12 (LexisNexis 2005) (stating that "the commission shall negotiate and enter into agreements or compacts with agencies of other states for cooperative regulatory efforts"); *In re* Exemption Application by Minn. Power for a 345/230 kV High Voltage Transmission Line Known as the Arrowhead Project, No. C4-01-1022, 2002 Minn. App. LEXIS 46, at *9 (Minn. Ct. App. Jan. 15, 2002) (considering reliability improvements to both Minnesota and Wisconsin citizens).

196. *See, e.g.*, 2003 BLACKOUT REPORT, *supra* note 67, at 32.

197. *See* MEYER & SEDANO, *supra* note 48, at E-16; Fox-Penner, *supra* note 49, at 14.

198. *See, Hearings*, *supra* note 56, at 13-16 (statement of David A. Svanda, President, Nat'l Ass'n of Regulatory Util. Comm'rs, and Comm'r, Mich. Pub. Svc. Comm'n); MEYER & SEDANO, *supra* note 48, at E-21; Dang, *supra* note 84, at 335; Fox-Penner, *supra* note 49, at 14; Ronald E. Russell, *Toward Federal/State Regulatory Harmony: Perspective of a State Regulator*, 9 CONN. J. INT'L L. 869, 875 (1994).

199. Jim Rossi, *Moving Public Law Out of the Deference Trap in Regulated Industries*, 40 WAKE FOREST L. REV. 617, 647 (2005).

200. Eisen, *supra* note 62, at 557.

1. NIMBYism and Environmental Concerns

Even if regional considerations are valid factors in a state's transmission siting decision process, local government bodies with siting authority are unlikely to weight them equally with local concerns.[201] Because local officials answer directly or indirectly to the community they serve, they are bound by law or prudence to give local concerns greater weight than they might objectively warrant:

> [E]ven when the regional or national benefits are well understood, facility siting opposition is almost universally driven by concern over the impacts on local residents. In effect, local residents are more often than not unwilling [to] accept what are perceived as significant impacts on their community and way of life in order to provide cost, reliability and environmental benefits to regional or national populations.[202]

This NIMBYism can prevent the approval of any new transmission line project, no matter how dramatic its benefits.[203] Even when the new infrastructure improves reliability and lowers prices, is necessary for national security, and replaces older, more heavily polluting facilities, siting attempts may fail.[204] These failures can occur in any jurisdiction and to any developer, regardless of the developer's competence and its efforts to satisfy local residents.[205]

The magnitude of the NIMBY problem has increased in recent years with the decrease in available land, the growing necessity of placing infrastructure close to existing populations, and the increasing levels and organization of local opposition.[206] These days:

> [T]he reasons for approval or denial of major energy facility infrastructure appears to have less to do with carrying out explicit national and state energy policies or with agency jurisdiction than with a combination of political support or opposition, choice of location, the strength and organization of local opposition, the quality of decisions and siting process execution on the part of project developers, regional market economics, and the rapidly changing nature of short-term reactionary politics.[207]

The emphasis on aesthetics and electromagnetic fields (EMF) in local siting decisions evidences this parochialism. Local outrage over aesthetics and EMF are a major factor in siting decisions and can derail vital transmission

---

201. HIBBARD, *supra* note 6, at 17.
202. *Id.*
203. *Id.* at 19.
204. *Id.*
205. *Id.*
206. *Id.* at 17.
207. *Id.* at 20.

expansions.[208] EMF are often an issue in certification proceedings, where opponents typically demand that the line be redesigned or rerouted, arguing that it unsafely exposes nearby residents to EMF.[209] EMF exposure also gives rise to personal injury claims, as well as inverse condemnation claims from nearby landowners, who assert that the new transmission line has reduced their property's value because of widespread public fear of EMF.[210]

In addition, many state regulatory bodies have responded to these concerns by developing EMF policies, many of which mandate "prudent avoidance."[211] Although not well defined, this standard is essentially a common-sense approach that EMF exposure should be avoided where it is inexpensive when balanced against the concerns of the citizenry.[212] Nevertheless, even ten years ago EMF expenses in the electric utility industry resulting from design and route changes, and from litigation, research, and regulatory compliance exceeded $1 billion annually.[213]

Although there is some evidence to the contrary, published studies have shown no consistent link between EMF levels and cancer incidence.[214] Nevertheless, fear of EMF and concern over the unsightliness of transmission lines has led to a very significant NIMBY effect.[215] Notably, the analogous federal statute regulating the siting of wireless telephone towers forbids localities from considering the EMF effects of facilities designed and constructed to Federal Communications Commission specifications.[216] Other

---

208. *See* San Diego Gas & Elec. Co. v. Super. Ct., 920 P.2d 669, 694 (Cal. 1996) (mentions EMF levels as a siting factor); City of Norwalk v. Conn. Siting Council, No. CV030524145S, 2004 WL 2361540, at *3 (Conn. Super. Ct. Aug. 18, 2004) (debating unsightliness at length in spite of great local need); Fl. Power Corp. v. State Siting Bd., 513 So. 2d 1341, 1343 (Fla. Dist. Ct. App. 1987) (denied because of EMF concerns); Barensfeld v. Pa. Pub. Util. Comm'n., 624 A.2d 809, 811 (Pa. Commw. Ct. 1993) (referring approvingly to an administrative judge's holding that EMF effects of a proposed line needed to be studied and could jeopardize the siting application); *In re* Narragansett Elec. Co., 544 A.2d 121, 126-27 (R.I. 1988) (holding scenic development of the region to be a major siting factor); Vt. Elec. Power Co. v. Bandel, 375 A.2d 975, 981 (Vt. 1977) (holding that the exercise of eminent domain for a transmission line would have "unduly interfere[d] with . . . scenic preservation"); Sierra Pac. Power, 64 C.P.U.C.2d 442, 468 (1996) (expecting up to 4% of total project cost to be spent on EMF mitigation and requiring a 300-foot right-of-way around the line to reduce EMF exposure).

209. Bogardus, *supra* note 114, at 1707.

210. *Id.* at 1708-09.

211. *Id.* at 1711-12.

212. *Id.* at 1712.

213. *Id.* at 1724-25.

214. *See, e.g.*, COMM. ON THE POSSIBLE EFFECTS OF ELECTROMAGNETIC FIELDS ON BIOLOGIC SYS., NAT'L RESEARCH COUNCIL, POSSIBLE HEALTH EFFECTS OF EXPOSURE TO RESIDENTIAL ELECTRIC AND MAGNETIC FIELDS 195 (1997), *available at* http://fermat.nap.edu/catalog/5155.html.

215. Dang, *supra* note 84, at 337; Rossi, note 200, at 647.

216. 47 U.S.C. § 332(c)(7)(B)(iv) (2000).

local environmental concerns may also be given undue weight.[217]

Although local opposition to new transmission lines stemming from environmental concerns is often dismissed as a NIMBY phenomenon (possibly as a result of the scientifically unjustified furor over EMF), these lines can have a significant impact on local environments.[218] "The mere fact that there is opposition to siting a particular facility does not necessarily signify parochialism or NIMBY at work."[219]

Some have suggested that the widespread opposition to the siting of new transmission facilities is grassroots energy-related activism that questions the social costs of deregulation and the rise of merchant power plants.[220] Most of these battles are, in fact, fought not by powerful national environmental organizations, but by coalitions of smaller local groups with names like, "A People's Energy Campaign," "Citizens for the Hudson Valley," and "Vermonters for Clean Energy."[221] This may be because organizations with a broader view often avoid disputes that seem to involve NIMBY-driven opposition to newer, cleaner, more reliable plants and facilities.[222] National environmental organizations are more concerned with "the global-warming and air-pollution effects of generation" than with localized aesthetic and health impacts of transmission.[223]

Although local groups may inordinately emphasize local land values and concerns, in some cases national groups may be missing an opportunity to press their agendas on significant policy issues.[224] It is difficult to separate the true policy wheat from the NIMBY chaff in a confrontation that is being played out in hundreds of small towns across the nation.

2. Parochialism and Corporate Favoritism

State and local businesses often wield a powerful influence over the political process that puts out-of-state competitors seeking to build interstate facilities at a disadvantage.[225] Local authorities tend to favor local businesses that hire local workers and provide local benefits. This preference may translate to political difficulties for competing businesses and projects, and

---

217. *See, e.g.*, Fla. Power Corp. v. State, Dep't of Envtl. Regulation, 638 So. 2d 545, 560-62 (Fla. Dist. Ct. App. 1994) (rejecting a proposal because of vague environmental concerns, including worries that maintenance might prevent herbaceous wetland from returning to forested wetland in 30-50 years).

218. Brown & Daniels, *supra* note 83, at 26; Chris Deisinger, *The Backlash Against Merchant Plants and the Need for a New Regulatory Model*, ELECTRICITY J., Dec. 2000, at 51, 51.

219. Brown & Daniels, *supra* note 83, at 26.

220. Deisinger, *supra* note 218, at 52-53.

221. *Id.*

222. *Id.* at 53.

223. *See* HIRST, *supra* note 8, at 11.

224. *See* Deisinger, *supra* note 218, at 53.

225. Brown & Daniels, *supra* note 83, at 24-25.

even to legislative action explicitly discouraging new investment. Several states, including Florida, have adopted moratoria on the construction of new merchant generators or have allowed them to be constructed only by in-state utilities.[226]

In *Tampa Electric Co. v. Garcia*,[227] the Supreme Court of Florida reviewed the decision to grant a determination of need for a proposed electric power plant.[228] The plant was to be a joint project between New Smyrna, a Florida public utility, and Duke Energy, a North Carolina-based, investor-owned wholesale utility company.[229] Under the proposal, the plant was to generate 514 megawatts, 30 megawatts of which Duke agreed to sell to New Smyrna at a discount rate.[230] Duke was to sell the remainder of the plant's capacity on the wholesale market, primarily within Florida.[231]

A determination of need was initially granted by the Florida Public Service Commission on the basis of an interpretation of "regulated electric companies" that included companies regulated by the FERC and not the state of Florida.[232] Several Florida public utility companies appealed the ruling.[233] The court reversed, finding that the statutory provision[234] allowed a determination of need to be granted only to those companies regulated by Florida law.[235] According to the court, the 1973 Florida legislation was intended to restrict the meaning of "need" to "demonstrated specified needs of . . . Florida customers" and to exclude "[t]he projected need of unspecified utilities throughout peninsular Florida" by rejecting wholesale utilities as applicants for determinations of need.[236] The court also dismissed an argument that the statute, as interpreted, was in violation of the Dormant Commerce Clause.[237] It noted that Congress explicitly left power plant siting and need determinations to the states in the 1992 Energy Policy Act.[238]

A dissent called this "a strained and artificial construction of various provisions of the legislative scheme that have little bearing on the issue before

---

226. Rossi, *supra* note 199, at 671; *Nervous of NOx*, *supra* note 97.

227. 767 So. 2d 428 (Fla. 2000).

228. *Id.* at 429.

229. *Id.* at 430.

230. *Id.*

231. *Id.*

232. *Id.* at 432.

233. *Id.* at 430.

234. Fla. Stat. Ann. § 403.519 (West 2002) (stating that "[o]n request by an applicant or on its own motion, the commission shall begin a proceeding to determine the need for an electrical power plant subject to the Florida Electrical Power Plant Siting Act").

235. *Tampa Elec.*, 767 So. 2d at 434. The court found that, under § 403.503(4), an applicant may be any "electric utility," and utilities are defined in § 403.503(13) as "regulated electric companies." *Id.* at 434.

236. *Id.* at 435-36.

237. *Id.* at 436.

238. *Id.*

us today."[239] The case is an example of a state government putting out-of-state competitors at a disadvantage by refusing to give siting approval even though the plan will provide power primarily within the state.

Considering the federal interest in competitive wholesale markets, such legislation may face substantial Dormant Commerce Clause challenges in the future. Recently, in *Granholm v. Heald*,[240] the U.S. Supreme Court struck down state laws that discriminated against out-of-state wineries in spite of the broad power accorded states by the 21st Amendment to regulate alcoholic beverages.[241] The Court noted that state laws violate the Commerce Clause if they mandate "differential treatment of in-state and out-of-state economic interests that benefits the former and burdens the latter."[242]

The Cross-Sound transmission line project between New York and Connecticut further illustrates the influence of local utilities over the approval of competitors' projects. A local company, serving mostly Connecticut residents, had a competing Cross-Sound line and sought to improve its own line rather than allow competition.[243] Although Connecticut cited environmental defects in delaying approval of the out-of-state project, some argued that the defects were minor and considered the Connecticut utility to have exercised significant influence.[244] Soon after the line was completed, Connecticut passed a moratorium on transmission lines across the Sound,[245] leaving the new line inoperable for more than a year.[246] Only a threatened regulatory decision on a related matter by the FERC brought the parties to the table to negotiate a resolution.[247]

Proponents of the existing siting system argue that only local authorities can take into account the myriad concerns of local citizenry.[248] They claim, for example, that unlike state and local governments, a federal committee located in Washington, D.C. would not be able to consider such factors as a transmission line's impact on historical monuments, park and recreational areas, and the local environment, including such details as the ecological balance of local wetlands and the viability of endangered species.[249]

Yet, it is not clear that a regional or national entity would not consider these issues. Surely the same local authority that currently makes siting

---

239. *Id.*

240. 125 S. Ct. 1885 (2005).

241. *Id.* at 1907.

242. *Id.* at 1895 (quoting Or. Waste Sys., Inc. v. Dep't of Envtl. Quality of Or., 511 U.S. 93, 99 (1994)).

243. *See* Rossi, *supra* note 199, at 646.

244. *Id.* at 645-46.

245. *Conn. governor signs moratorium on grid projects, keeping Cross Sound in limbo*, POWER MARKETS WK., June 30, 2003, at 31.

246. Rossi, *supra* note 199, at 646.

247. *See* Bruce Lambert, *New York and Connecticut Agree to End Cable Dispute*, N.Y. TIMES, June 25, 2004, at B6.

248. MEYER & SEDANO, *supra* note 48, at E-21.

249. *See* Dang, *supra* note 84, at 337-38.

decisions, or at least a similar one, would assist a new decision-making body by advising it of local concerns. The fear may be that a regional entity would dismiss these concerns. However, states setting up their own regional process would certainly ensure its responsiveness to local concerns. Such a process would likely be sensitive to local politics even if it takes a broader view and is allowed to consider interstate benefits in approving new projects.[250] Even executive agencies such as the FERC are in some way responsive to constituents. National organizations and interests, including lobbying groups for local and state governments, can wield considerable political influence. In addition, federal backstop authority would likely encourage negotiated agreements between the parties and would rarely have to be used, given even the current effectiveness of FERC pressure.[251]

In any event, the likelihood of bias seems far greater for siting decisions made at local levels. Local governments may introduce significant externalities to the siting process, thus giving inordinate weight to local factors at the expense of nationwide benefits to consumers.[252]

E. Interstate Disagreements

Utilities seeking to construct new interstate transmission facilities also face a lack of coordination between the various state and federal agencies from which they must obtain approval.[253] Each jurisdiction and agency has its own independent proposal approval process.[254]

> Interstate transmission siting approvals are among the most complex and lengthy permitting processes anywhere. . . . [M]any counties or localities along the path of the line require local approvals, and these communities have an important political voice in the state approval process as well. If the line crosses any federal lands or waterways, a number of federal approvals may be needed from a disparate set of agencies. In some cases, permits must be applied for and received in sequence.[255]

Because each state and agency has its own priorities and is not necessarily concerned with the effects of the proposal on the other parties involved, conflicts can develop during this process.[256] One function of the approval process is to analyze alternatives to the proposal and to determine the method

---

250. MEYER & SEDANO, *supra* note 48, at E-21.

251. *See* Rossi, *supra* note 199, at 670.

252. *See* Carl J. Levesque, *Regulators' Forum: Can FERC and States Unite?*, PUB. UTIL. FORT., Nov. 15, 2001, at 14, 20 (statement of Thomas Welch, Chairman, Maine Public Util. Comm'n).

253. *See* MEYER & SEDANO, *supra* note 48, at E-16, E-18.

254. Fox-Penner, *supra* note 49, at 13.

255. *Id.*

256. MEYER & SEDANO, *supra* note 48, at E-8, E-45; MUCHOW & MOGEL, supra note 93, § 81.01(5).

that provides the most benefit at the lowest cost.[257] If each state or agency does not consider the ways in which the various proposals effect all the parties involved, they are likely to come to disparate conclusions regarding the benefits and costs of various alternatives. When one state approves the initial proposal, and then a different state or federal agency requires changes to the design, the initial state must approve these changes.[258] In the end, the utility must satisfy each party involved in the siting process, an undertaking that can involve considerable time and expense due to the lack of coordination between parties.

Parochial local concerns can amplify the problem of obtaining siting approval from multiple states. If a proposed project would transmit power from an area with plentiful electric power to an area where power is more expensive, and the transmission line would have to pass through a third state, obtaining approval from the third state would be nearly impossible.[259] Courts have routinely held that transmission lines running through a state, but not providing electricity directly to that state's citizens, provide no direct benefit to the state and cannot be granted the use of that state's power of eminent domain.[260] Similarly, political grandstanding in one locality could delay an entire multi-state electric transmission facility project because all parties must approve a single proposal before construction can begin.

While there are many examples of multi-state transmission projects facing delays and expense,[261] there are undoubtedly even more instances where new construction was never pursued because of expected approval problems.[262] When deciding whether to pursue the construction of a new transmission line crossing several states and passing various federal lands, a utility must weigh the costs and benefits of the new line, including the cost of obtaining the necessary approval and the possibility that approval will be denied. The increased cost of negotiating approval independently with all of these various interests may deter the construction of new transmission facilities when compared with a single, streamlined one-stop approval process of lesser expense. Reports show that new transmission construction is not keeping up with the growing demand.[263]

---

257. MEYER & SEDANO, *supra* note 48, at E-5, E-32.

258. *Id.* at E-37.

259. Carl J. Levesque, *Stringing Transmission Lines, Untangling Red Tape*, PUB. UTIL. FORT., Sept. 1, 2001, at 46, 50 (statement of Tim Gallagher, manager of technical services at FERC).

260. *See Clark*, 198 So. 2d at 371 (holding that a one-way transmission line from Florida into Georgia did not constitute a public use); *Conerly*, 460 So. 2d at 109-13 (finding a transmission line solely for transferring power out of state not to be a public use).

261. MEYER & SEDANO, *supra* note 48, at E-9, E-15, E-19, E-38.

262. *Id.* at E-12 to 13; Brown & Daniels, *supra* note 83, at 26.

263. MEYER & SEDANO, *supra* note 48, at E-1.

III. THE ENERGY POLICY ACT OF 2005 AND TRANSMISSION SITING OVERSIGHT

A. Background

The Energy Policy Act of 2005[264] contains, in addition to its better-known features,[265] an important new section on the siting of electric transmission lines. Proposals to reform the process of electric transmission facility siting are not new. Industry experts have been pushing for a greater role for the FERC for over a decade.[266] Until recently, those efforts met with little success. However, in the aftermath of the September 11, 2001, terrorist attack, the Bush administration emphasized the need to reconsider energy policy.[267] Lawmakers have been encouraged to overcome the difficulty of dealing with the many special interests in the energy industry.[268]

The Act's passage was highly criticized,[269] largely by environmental groups.[270] The Sierra Club declared that "this bill funnels billions of taxpayer dollars to polluting energy industries, and opens up our coastlines and wildlands to destructive oil and gas activities."[271]

B. General Provisions of the New Act

According to the Conference Report preamble, the purpose of the Act is, "to ensure jobs for our future with secure, affordable, and reliable energy."[272] The principle purposes of the Act are:

---

264. 2005 Act, *supra* note 1.

265. *See* discussion *infra*, Part III.B.

266. *See* Rossi, *supra* note 199, at 668.

267. *See id.* at 669.

268. *Id.* at 668-69 (discussing the political problems in getting a comprehensive energy bill passed, including collective action, earmarks, legislative pork, and agenda setting problems).

269. *See, e.g.*, Opinion, *Big Oil's, Big Break; Bill Both Insulates, Rewards Thriving Industry*, SAN DIEGO UNION-TRIBUNE, Aug. 10, 2005, at B-8; Peter Z. Grossman, Editorial, *If you Like Pork, You'll Love the New Energy Bill*, INDIANAPOLIS STAR, Aug. 9, 2005, at 8A; Michael Grunwald and Juliet Eilperin, *Energy Bill Raises Fears About Pollution, Fraud; Critics Point to Perks for Industry*, WASH. POST, July 30, 2005, at A01.

270. *See* Mike Soraghan, *No Drilling Rev-Up in Rockies Bush's Signature on Energy Bill Eases Permit Process, But Industry Already Going as Fast as It Can, Faster Approval for Gas Wells is Among the Law's Regulatory Changes and Energy Initiatives.*, DENVER POST, Aug. 9, 2005, at A-01; Richard W. Stevenson, *Bush Signs an Energy Bill That Had Been a Longtime Priority*, N.Y. TIMES, Aug. 9, 2005, at A12; SUMMARY OF THE HARMFUL PROVISIONS IN THE ENERGY BILL (July 28, 2005), *available at* http://www.sierraclub.org/globalwarming/bush_plan/energybill_bad_provisions7_28_05.pdf.

271. Press Release, Carl Pope, Executive Director, Sierra Club, America Needs Real Energy Solutions (July 27, 2005), *available at* http://www.sierraclub.org/pressroom/ releases/pr2005-07-28.asp.

272. H.R. REP. NO. 109-190, at 1 (2005) (Conf. Rep.).

- To provide incentives for energy conservation, including new energy efficiency standards for appliances and renewables such as hydroelectric facilities and ethanol.

- To provide billions of dollars for research and development into new energy technologies, including clean coal and hydrogen.

- To improve the reliability of the transmission grid and the efficiency of the electricity market through the establishment of an independent organization to enforce mandatory reliability standards and the expansion of FERC authority to ensure open access to transmission lines and to punish market manipulation.

- To encourage the development of nuclear power through new loan and insurance programs.[273]

C. Act Section 1221 — Siting of Interstate Electric Transmission Facilities

In *New York v. Federal Energy Regulatory Commission*,[274] the United States Supreme Court recognized that the Interstate Commerce Clause[275] gives the federal government the right to assert jurisdiction over the entire transmission grid, "all the way to a consumer's toaster."[276] This decision was consistent with the Supreme Court's well known 1942 holding in *Wickard v. Fillburn*.[277] Recently, in *Gonzales v. Raich*,[278] the Court declared that, "[w]hen Congress decides that the 'total incidence' of a practice poses a threat to a national market, it may regulate the entire class."[279] The Court stated:

> Our case law firmly establishes Congress' power to regulate purely local

---

273. ICF CONSULTING, ICF CONSULTING IDENTIFIES THE U.S. ENERGY ACT'S IMPLICATIONS FOR THE GLOBAL ENERGY SECTOR (2005), *available at* http://www.icfci.com/Newsroom/energy-bill-2005.asp; COMM. ON ENERGY & COMMERCE DEMOCRATIC STAFF, SUMMARY OF POLICY PROVISIONS OF THE ENERGY POLICY ACT OF 2005 CONFERENCE REPORT (2005), *available at* http://www.house.gov/commerce_democrats/energy/Energy_highlights.pdf.

274. 535 U.S. 1, 23-24 (2002).

275. U.S. CONST. art. I, § 8, cl. 3 ("The Congress shall have Power . . . To regulate Commerce with foreign Nations, and among the several States . . . .").

276. Eisen, *supra* note 62, at 572.

277. Wickard v. Filburn, 317 U.S. 111, 127-29 (1942) (holding that the production of wheat for one's own personal consumption affects interstate commerce and can be regulated by the federal government).

278. 125 S.Ct. 2195, 2201 (2005) (upholding under the Commerce Clause federal regulation of marijuana grown and used non-commercially within California for medical purposes).

279. *Id.* at 2206 (quoting Perez v. United States, 402 U.S. 146, 154-55 (1971)).

> activities that are part of an economic "class of activities" that have a substantial effect on interstate commerce. As we stated in *Wickard*, "even if appellee's activity be local and though it may not be regarded as commerce, it may still, whatever its nature, be reached by Congress if it exerts a substantial economic effect on interstate commerce."[280]

Extending this logic to transmission lines, it is clear that a transmission line, no matter how local, has some effect on the supply of electricity and in turn effects the market price of electricity. Therefore, the federal government must have the authority to regulate the siting and construction of electric transmission lines.

1. The Problem of Federal vs. State Jurisdiction Over Transmission

Since 1935, Section 201 of the Federal Power Act ("FPA")[281] has given the Federal Power Commission, and now the FERC, broad authority over "transmission of electric energy in interstate commerce and the sale of such energy at wholesale in interstate commerce," including rate setting and regulation of terms and conditions of service.[282] However, the FPA also states that federal regulation "extend[s] only to those matters which are not subject to regulation by the States."[283] "[S]tates have historically exercised exclusive jurisdiction over transmission siting."[284] Therefore, the FPA preserves that jurisdiction and transfers no transmission siting authority to the FERC.[285]

In addition, the FPA specifically reserves jurisdiction for the states over facilities that generate electricity and transmit it locally, transmit it in intrastate commerce, or consume it themselves.[286] Courts have interpreted this section as intended to preserve state "regulatory control of purely state-wide facilities," even those that affect interstate commerce (as all do to some extent).[287] These limitations on the power of the federal regulatory commission are a clear indication that Congress intended that the FPA maintain "the existing scope of state authority."[288] The FPA was enacted only "to fill the gap in state regulatory authority that the Supreme Court had created by holding that no state could regulate interstate wholesales of electricity."[289]

Therefore, under the Federal Power Act, the FERC does not have the power to regulate purely intrastate generating facilities or transmission siting.

---

280. *Id.* at 2205-06 (quoting *Wickard*, 317 U.S. at 125) (citing *Wickard*, at 128-29; *Perez*, 402 U.S., at 151) (citations omitted).

281. 16 U.S.C. § 824(a) (2000).

282. *Id.*

283. *Id.*

284. *See* Dang, *supra* note 84, at 336.

285. *Id.*

286. 16 U.S.C. § 824(b)(1) (2000).

287. Mateer, *supra* note 14, at 781.

288. *See* Dang, *supra* note 84, at 336.

289. Pierce, *supra* note 10, at 466.

However, the FERC's control over transmission transactions is near complete:

> [T]he Supreme Court has interpreted the FERC's exclusive and non-delegable jurisdiction over interstate transactions to include intrastate wholesale transactions on transmission lines connected to an interstate grid. If the electric power conceivably flows in interstate commerce, the FERC may assert jurisdiction. Even when both of the contracting parties and the electrical pathway between them are within one state, if the system is interconnected and capable of transmitting power across a state boundary, the FERC has asserted jurisdiction. Because transmission lines in all states except Alaska and Hawaii are connected to an interstate grid, there is often little opportunity for state regulation of transmission transactions.[290]

Some see this dichotomy of regulatory authority as a major obstacle to the successful restructuring of the electricity market.[291] "[T]he vision of a competitive wholesale market is being advanced by federal authorities who lack siting jurisdiction, while siting authorities may well lack statutory authority with the same vision."[292]

The 2005 Energy Policy Act's siting provisions address this issue. The Act provides for federal backstop siting authority for transmission projects in critical regions where one of the states involved is not authorized to consider regional benefits. These provisions are intended to streamline the process of siting critical regional transmission lines and facilities, thus ensuring that the national electricity transmission grid has adequate capacity and increased reliability. The Act should also make it easier for various agencies and jurisdictions to coordinate their efforts and avoid conflicting decisions. It also provides strong incentives for states to work together in regional organizations. Any analysis of the Act's siting provisions must consider whether they will achieve the objective of increased national transmission grid capacity and reliability.

2. National Interest Electric Congestion Corridors and Federal Siting Authority

The Energy Policy Act of 2005 adds several important siting changes to the Federal Power Act.[293] First, it institutes a study of electric transmission congestion, to be conducted every three years by the Secretary of Energy ("Secretary") in consultation with affected states.[294] The Secretary, upon reviewing this study, is to designate geographic areas experiencing congestion as "national interest electric transmission corridor[s]" if such a designation would benefit national defense, energy independence or policy, or regional

290. MUCHOW & MOGEL, *supra* note 93, at § 81.04(3) (citations omitted).
291. *See, e.g.*, Pierce, *supra* note 10, at 466; Brown & Daniels, *supra* note 83, at 24.
292. Brown & Daniels, *supra* note 83, at 24.
293. 16 U.S.C. § 824 et seq. (2000).
294. 2005 Act, *supra* note 1, § 216(a).

economic development.[295] The FERC could then issue permits for the construction of transmission facilities in such areas if an affected state is not authorized to consider interstate benefits, the permit applicant cannot qualify for a permit because it does not serve end-users in the state, or the state siting authority has unnecessarily delayed or conditioned permit approval.[296] For a project to get FERC permitting approval it would also have to be interstate in nature, be consistent with public interest, significantly reduce transmission congestion, and maximize the use of existing facilities to minimize its aesthetic and environmental impact.[297] The bill is careful to note that all affected parties will have an opportunity to present their view of any proposal that is before the FERC for permitting.[298]

The provisions also give the federal permit holder the right to use the federal power of eminent domain to obtain necessary easements for which negotiations have failed, and to conduct eminent domain proceedings in federal court.[299] In such cases, federal courts are required, when possible, to adopt the procedure used by the state court where the property is located.[300] When the federal power of eminent domain is exercised, the utility must compensate the owner for the value of the property taken, and for any resulting loss in value of the owner's remaining property.[301]

3. Streamlined Federal Authorization and Environmental Review

Further, the additions to the FPA allow for streamlined, one-stop federal authorization of all transmission projects and for coordination with other entities whose authorization is necessary.[302] The Department of Energy is to "act as the lead agency . . . [in] coordinating all . . . [f]ederal authorizations and . . . environmental reviews" of transmission facility proposals.[303] The Secretary will coordinate this federal authorization process with other necessary parties and set binding milestones and deadlines for the federal process and for any other entity that is willing to combine its efforts in this respect.[304] In addition, "the Secretary, in [conjunction] with . . . affected agencies, . . . [is to] prepare a single environmental review document, . . . [to serve] as the basis for all [federal] decisions" regarding the proposal.[305] The goal is timely and efficient decisions; the federal authorization must be

295. *Id.* § 216(a)(4).
296. *Id.* § 216(b)(1).
297. *Id.* § 216(b)(2)-(6).
298. *Id.* § 216(d).
299. *Id.* § 216(e)(1).
300. *Id.* § 216(e)(3).
301. *Id.* § 216(f)(2).
302. *Id.* § 216(h)(1)-(9).
303. *Id.* § 216(h)(2).
304. *Id.* § 216(h)(3)-(4).
305. *Id.* § 216(h)(5)(A).

completed within one year unless preempted by another federal law.[306]

A 60-day pre-approval process will allow prospective applicants to determine both the likelihood of obtaining approval from the necessary agencies and the key issues involved, without subjecting themselves to the full expense of the permitting process.[307] In cases of a necessary federal authorization denial or a missed deadline, the provisions authorize a presidential review.[308] Upon the filing of an appeal, the President must decide within 90 days whether to issue the authorization or to deny the application.[309] The Secretary is to consult regularly with the FERC and with reliability and transmission organizations approved by the FERC in carrying out these federal authorizations.[310]

The Act also provides for the implementation of the streamlined federal authorization process. The Secretary is to "issue any regulations necessary to implement" the Act within 18 months of the Act's passage.[311] Within one year of passage, the Secretary and the heads of all the Federal agencies with authorization authority are to "enter into a memorandum of understanding to ensure" streamlined permitting and environmental review processes.[312] Multistate entities, Indian tribes, and State agencies may also join this memorandum.[313] In addition, the head of each agency with authorization authority is to allocate funds and designate a senior official for implementation of the Secretary's regulations and the memorandum of understanding.[314]

4. Regional Transmission Siting Agencies

The new provisions also support the formation of regional transmission siting agencies. One addition allows three or more contiguous states to form such an agency for the purpose of carrying out those states' transmission facility siting responsibilities.[315] The regional agency will have permitting authority for the siting of transmission facilities, even in national interest electric transmission corridors.[316] The commission has no authority to issue a permit for a facility within a participating state unless it is on federal land or the states in the agency are in disagreement.[317] The proposal must also meet the other requirements for federal permitting.[318]

---

306. *Id.* § 216(h)(4)(B).
307. *Id.* § 216(h)(4)(C).
308. *Id.* § 216(h)(6)(A).
309. *Id.* § 216(h)(6)(B)-(C).
310. *Id.* § 216(h)(9).
311. *Id.* § 216(h)(7)(A).
312. *Id.* § 216(h)(7)(B)(i).
313. *Id.* § 216(h)(7)(B)(ii).
314. *Id.* § 216(h)(7)(C).
315. *Id.* § 216(i)(1).
316. *Id.* § 216(i)(3).
317. *Id.* § 216(i)(4).
318. *Id.* § 216(i)(3)-(4).

D. Analysis

1. Federal Permitting

At first glance, the Act's new federal siting authority appears to have the potential to fix several of the siting problems currently plaguing the industry. It allows the federal government to step in and approve the construction of new electric transmission facilities where a particular state involved is not authorized to consider interstate benefits, is discriminating against an out-of-state company, or is otherwise unnecessarily delaying or conditioning permit approval.[319] Those situations all arise frequently and are currently creating approval difficulties for important transmission facilities. However, there are limitations on the use of this newly granted federal power. The power can only be used in an area identified as a "national interest electric transmission corridor" ("Transmission Corridor").[320] To be federally permitted, the project must be interstate in nature and "significantly reduce transmission congestion."[321] It is uncertain exactly how liberal the Secretary and the commission will be in designating such corridors and in interpreting those project requirements. A Transmission Corridor can be designated anywhere that would benefit regional economic development.[322] That language may give the Secretary great flexibility when designating Transmission Corridors. Although these limitations will likely allow some important projects to slip through the cracks, they do not seem to undermine the significance of the new federal authorization powers.

While these provisions make it easier for critical interstate transmission facilities to obtain siting approval, it is important to note that permitting and the use of eminent domain are separate issues. The power of eminent domain does not immediately follow the granting of a siting permit. The Act provides for the use of federal eminent domain power, but only for projects that are federally permitted.[323] Therefore, it is still possible that a proposal that is granted a state siting permit will be denied the use of eminent domain power. The Act does nothing to resolve that issue.

The Act also allows states to avoid the possibility of federal pre-emption in transmission siting if they cede their siting authority to a regional entity. In that case, federal siting authority would be limited to situations where the members of that entity are in disagreement.[324] The threat of federal backstop siting authority will undoubtedly encourage states to form regional entities with siting authority so as to avoid federal pre-emption, which may be the

319. *Id.* § 216(b)(1).
320. *Id.* § 216(b).
321. *Id.* § 216(b)(2)-(6).
322. *Id.* § 216(a)(4).
323. *Id.* § 216(e).
324. *Id.* § 216(i)(4).

main purpose behind the federal permitting provisions. In any case, this does not reduce the Act's impact. The permitting provisions were probably created to avoid a situation where states fail to internalize all of a project's benefits in making a siting decision or discriminate against out-of-state companies. These are concerns that would most likely disappear if siting authority were vested within a regional agency, depending somewhat upon the agency's composition.

2. Streamlined Authorization and Cooperation

Having implemented federal permitting provisions to deal with state permitting problems and interstate discrimination, the Act attempts to resolve interstate disagreements and reduce the length, complexity, and expense of the siting process with its new streamlined federal authorization process. Requiring all the affected federal agencies to work in concert to create a single environmental review should reduce conflict and confusion in the approval process. Mandating binding milestones and requiring all federal authorizations to be completed within a year should help reduce the length of the process. The pre-approval process also seems likely to reduce expense and delay, as it will quickly bring to light any major problems that an applicant needs to address. Coordinating all required federal authorizations under one lead agency seems to be a very salutary development, and these provisions, in combination with new regional siting agencies, may actually reduce time and expense.

The Act's implementation provisions, however, are somewhat vague. With all of the new deadlines and procedures, it is important that agencies are still able to make appropriate, well-considered decisions about whether to authorize new transmission proposals. Therefore, the manner in which these streamlining provisions are implemented will be critical. The Act simply requires the Secretary to advance all necessary regulations within 18 months,[325] for agencies involved to enter a memorandum of understanding within 12 months,[326] and then for those agencies to devote necessary resources to successfully implement the regulations and memorandum. Much will depend on those regulations and their implementation.

IV. PROPOSED SOLUTIONS

Will the 2005 Energy Policy Act mark the beginning of a new era of transparent, efficient, and effective transmission siting regulation, or is it too limited and vague to succeed? Will the Act help overcome the barrier of parochialism and encourage renewed investment in critical transmission infrastructure, or will the transmission shortage crisis continue to worsen? Attempts to reform the siting process for wireless telecommunications towers

325. *Id.* § 216(h)(7)(A).
326. *Id.* § 216(h)(7)(B)(i).

have demonstrated the ineffectiveness of overly deferential, half-hearted measures providing process-based rather than substantive standards.[327] It is too early to say for sure what the results will be, but it is apparent that while the 2005 Act is a step in the right direction, some significant transmission siting difficulties will remain. The Act adequately addresses some difficulties, but others may require further legislation or regulation in order to alleviate the existing transmission shortage.

A. Regional Benefits in Siting and Eminent Domain

Many experts have called for a shift of transmission facility siting authority from states and localities to regional or national entities in order to appropriately account for regional benefits in the siting process.[328] State and local governments and organizations largely oppose this shift, asserting that adjustments made under the existing siting regime are sufficient to deal with any transmission problems and that such a move would trivialize the concerns of local communities.[329]

Because a regional siting authority would eliminate externalities in the approval process by considering benefits over an interstate area, some local concerns will not be weighted as heavily as they are by sometimes-parochial local governments. It is understandable that local organizations wish to maintain their influence and therefore oppose the proposed measures. Maintaining inefficiencies, however, is not a compelling reason to keep the current system. Although some states have made progress in their permitting processes, many still do not weigh interstate benefits appropriately.[330] With a growing national shortage of transmission infrastructure, the cost of allowing the states to slowly adjust their siting laws is probably greater than the cost of transitioning to a regional or federal siting regime. The new Act addresses this concern by giving the FERC siting authority over interstate transmission projects proposed in Transmission Corridors when a state cannot account for interstate benefits in its approval process.[331]

It may seem that this provision will allow the siting of critical transmission facilities where it would otherwise be impossible. However, states accord varying weight to regional benefits. While many states do not flatly deny the importance of regional benefits, they nevertheless give such benefits short

---

327. *See generally*, Steven J. Eagle, *Wireless Telecommunications, Infrastructure Security, and the NIMBY Problem*, 54 CATH. U. L. REV. 445 (2005) (referring to the "National Wireless Telecommunications Siting Policy" contained within the Telecommunications Act of 1996, 47 U.S.C. § 332(c)(7) (2000), imposing procedural, but generally not substantive, federal standards for siting wireless telecommunications towers).

328. *See* HIRST, *supra* note 8, at 21; TAPS, *supra* note 117, at 21; Brown & Daniels, *supra* note 83, at 34; Dang, *supra* note 84, at 349; Fox-Penner, *supra* note 49, at 13.

329. *See supra* discussion Part II.D.

330. *See supra* discussion Part II.C.

331. 2005 Act, *supra* note 1, § 216(b)(1)(A)(ii).

shrift when weighing them against costs within the state.[332] States that do explicitly reject regional benefits will likely liberalize their legislation slightly rather than cede siting authority to the federal government, as states have generally opposed any loss of such authority. Thus, this provision is not in itself an answer to the parochialism inherent in the siting approval processes in many states.

Congress also granted siting power to the FERC in Transmission Corridors where the state authority has withheld approval for more than a year or conditioned its approval in such a way as to rob the project of its ability to relieve grid congestion.[333] This provision should ensure that critical transmission proposals are approved in a reasonable amount of time, relieving one major source of added expense and uncertainty in the industry.[334]

The Act states specifically that the FERC may issue a siting permit for a proposed transmission project within a Transmission Corridor where a state has "withheld approval for more than 1 year after the filing of an application seeking approval pursuant to applicable law."[335] Can the FERC intervene only where the state has failed to act altogether, or can it intervene when the state has explicitly denied approval? Neither the Act nor the commentary answer this very important question. If a state can deny an application within one year to avoid federal jurisdiction, it could completely circumvent the Act by denying an application right before the deadline and allowing another application in the future. Congress did not use the term, "denied approval," but rather used the term, "withheld approval."[336] In the past, these terms have been distinguished from one another in a public utilities context.[337]

Neither legislative hearings nor legal commentary resolve this matter. The issue has generally been ignored, but a few commentators have described the federal backstop authority as being available only when a state "fails to act" within a year,[338] or stated that this authority "provide[s] incentives for reaching

---

332. *See supra* discussion Part II.C.

333. 2005 Act, *supra* note 1, § 216(b)(1)(C)(i)-(ii).

334. *See supra* discussion Part II.A.

335. 2005 Act, *supra* note 1, § 216(b)(1)(C)(i).

336. Id.

337. *See* Cent. Kan. Elec. Coop. Ass'n v. State Corp. Comm'n, 196 P.2d 212, 214 (Kan. 1948); Pub. Serv. Co. of N.M., 211 P.U.R.4th 169 (N.M.P.S.C. 2001) ("[C]ondition 15(e) is rewritten as follows: 'PNM must agree that it will not challenge the Commission's authority to withhold approval of the Western merger. PNM can only challenge a denial of approval based on the merits.").

338. *See* STEPHEN ANGLE ET AL., VINSON & ELKINS, IMPLICATIONS OF THE ENERGY POLICY ACT OF 2005 2 (2005) *available at*, http://www.vinson-elkins.com/pdf/resources/EPAHandout081805.pdf; *see also Hearings*, *supra* note 56, at 36, 213 (stating that "having a 'fallback' right of eminent domain, as laid out in the House draft . . . would provide a motivation to ensure that state inaction does not occur" and suggesting "[f]ederal backstop transmission line siting authority . . . where states have failed to act") (statements of John Anderson, Executive Director, Electricity Consumers Resource Council & Dr. Gregory Reed, Vice President, National Electrical Manufacturer's Association).

decisions on siting applications within a year of filing."[339] Others, however, have asserted that this new authority allows the FERC to override state entities and issue permits where the state has denied approval.[340] The latter is the more likely interpretation, as the former interpretation would eviscerate the provision. The term "withhold approval" probably means the failure to act or the outright denial of approval.

The fact that federal permitting exists at all may increase the number of permit applications denied by state and local boards. Some have speculated that a state utility commission, faced with a justified but unpopular proposal for a new transmission facility, may "pass the buck," allowing the federal government to permit the facility and thereby avoid a political backlash.[341] Such a trend would further increase permitting delays and complexity.

In addition, the Act allows use of the federal power of eminent domain, but only for those projects that receive federal permits.[342] However, eminent domain and siting approval are two distinct processes.[343] A project that has obtained state siting approval may still be unable to use the state power of eminent domain. The Act also allows states to avoid the threat of federal preemption by joining a regional agency with siting authority (except in the case of disagreement among the members of the compact).[344] While encouraging regional rather than state-based siting, this provision again ignores the necessity of eminent domain for the construction of electric transmission lines. A regional siting authority probably would not be vested with state eminent domain powers, and even if states did agree to relinquish their powers to a regional agency, state courts could find such a delegation unconstitutional.[345]

---

339. SKADDEN, ANALYSIS OF THE ENERGY POLICY ACT OF 2005 28 (2005), *available at* http://www.skadden.com/content/Publications/Publications1065_0.pdf.

340. *See Hearings*, *supra* note 56, at 16 (stating that "the FERC will have authority to override state decision processes on transmission siting, if that state is not in an RESC") (statement of David A. Svanda, President, Nat'l Ass'n of Regulatory Util. Comm'rs (NARUC)); GARDERE, MEMORANDUM RE: ENERGY POLICY ACT OF 2005 3 (2005) (stating, "when state commissions do not provide approvals . . . FERC is granted limited authority to issue construction permits"), *available at* http://www.gardere.com/Content/hubbard/tbl_s31 Publications/FileUpload137/1280/Domenici-Barton%20Memo.pdf; LEBOEUF LAMB, REPORT ON THE DOMENICI-BARTON ENERGY POLICY ACT OF 2005 6 (2005) (suggesting that "if the State commissions do not approve a project . . . the Act gives FERC authority"), *available at* http://www.llgm.com/files/Publication/4a8b8c70-5614-4247-b45f-1f37cc86165c /Presentation/PublicationAttachment/cba582dc-202c-46ae-ba3f-232daae48db2 /article_1129.pdf; SIMPSON THACHER, ENERGY POLICY ACT OF 2005: IMPLICATIONS FOR THE ELECTRIC UTILITY INDUSTRY 5 (2005) (stating that the "FERC will be able to override state authorities"), *available at* http://www.stblaw.com/content/ publications/pub519.pdf.

341. *See* MEYER & SEDANO, *supra* note 48, at E-21.

342. 2005 Act, *supra* note 1, § 216(e).

343. Brown & Daniels, *supra* note 83, at 24.

344. 2005 Act, *supra* note 1, § 216(i)(4).

345. At least one court has stated that the power of eminent domain can only be used for the benefit of state citizens. *Conerly*, 460 So. 2d at 113; *see supra* notes 120-27 and

Congress must ratify interstate compacts (which would be necessary to create a Regional Transmission Siting Agency) and could imbue regional siting agencies with federal eminent domain powers. However, there is no such provision in the Act, making it impossible to know whether Congress will give regional siting agencies this power or how the power would be constrained.

The Act, in its current form, does little to defeat the obstacle posed by some state courts that are unwilling to consider regional benefits in eminent domain proceedings. The current act would be improved with a federal eminent domain provision parallel to the federal siting provision. The federal power of eminent domain should be available, at least in Congestion Corridors, where state courts are unable to consider regional benefits for eminent domain purposes. Eminent domain power held at the regional level would likely accomplish the same objective with less bureaucracy. However, granting eminent domain power to regional entities would have to be handled carefully to avoid court challenges.

The backstop federal siting authority is a significant step in the right direction. It pressures states to give some weight to regional benefits and to promptly approve proposals. If the Act is indeed interpreted as allowing the FERC to overturn state denials of siting approval in Transmission Corridors, states will have a powerful incentive to either approve these projects or to cede their siting authority to a regional body. If the Act is not so interpreted, it will be weak and ineffectual, providing little incentive for states to change their regulatory behavior. In any case, the Act's efficacy is strongly dependent on how it is implemented. How the Secretary will use the power to designate Transmission Corridors and what deadlines will be set for the FERC permitting process are two important questions the Act leaves open.

A better solution would be to simply mandate regional siting. Then there would be no concerns about the precise delineation of Transmission Corridors, states "passing the buck" to the federal government on siting decisions, or regional benefits ever being given less than appropriate consideration. In addition, evaluating the cost-effectiveness of various transmission investments would be easier with all siting power at the regional level rather than under the system of bifurcated responsibility the Act would create.[346] Faced with the stringent opposition of state organizations and officials, Congress has enacted provisions with vague wording and uncertain procedures for implementation, and their effectiveness will not be apparent for years to come.[347]

---

accompanying text.

346. *See Hearings*, *supra* note 56, at 26 (statement of Gerald Norlander, Chairman, Electricity Committee, National Association of State Utility Consumer Advocates (NASUCA)).

347. Congress did something similar when it enacted new siting provisions for telecommunications towers in the Telecommunications Act of 1996. *See* Eagle, *supra* note 327, at 491, 495.

B. The NIMBY Phenomenon and State Parochialism

NIMBY opposition is an increasingly large factor in the siting of electric transmission lines.[348] The source of this opposition derives from the decline in available land and the proximity of necessary infrastructure to residential populations, so the phenomenon is likely to continue to grow.

Nevertheless, it is possible to minimize the impact of NIMBY opposition, and of parochialism in general, on the successful siting of transmission capacity expansion projects. One suggestion for dealing with NIMBYism is a shift to federal siting authority, as federal officials are less likely to be influenced by parochial concerns than the local siting councils prevalent in the existing siting regimes.[349] Siting agencies could also educate developers about potential pitfalls to ensure that proposals minimize the perceived affront to local residents.[350]

NIMBYism is a problem when small costs to a small group of vocal residents prevent the siting of infrastructure that provides large benefits to a large group of people. Such inefficiency results only when concerns of negatively affected residents are given disproportionate weight. A regional or federal siting authority would be less likely to bow to narrow local interests. The federal siting authority that the 2005 Act provides is, however, somewhat limited, while the NIMBY problem is widespread.

Federal permitting under the Act is possible only where the state cannot consider regional benefits or withholds its approval for more than a year. Regional siting is encouraged, but whether states will actually cede their siting authority to regional agencies is unknown. The Act does have a provision designed to prevent parochial corporate favoritism. Federal permitting is also triggered if a utility does not qualify to apply for a siting permit because the applicant does not serve end-use customers in the state.[351] That provision, however, addresses only one small aspect of the parochialism that pervades the current siting regime, albeit a particularly pernicious one. Universal regional siting is a better solution for parochialism just as it is for the consideration of regional benefits in the approval process.

The Act also provides for streamlined one-stop federal authorization of transmission facilities with binding milestones and deadlines.[352] All federal agencies are required to combine their authorizations into this new one-stop authorization process, and states and regional entities are also encouraged to coordinate their efforts with this new process.[353] The Act provides for an associated pre-approval process that identifies the key issues a proposal will

348. *See supra* discussion Part II.D.1.
349. *See* HIBBARD, *supra* note 6, at 17.
350. *Id.* at 22.
351. 2005 Act, *supra* note 1, § 216(b)(1)(B).
352. *Id.* § 216(h)(3).
353. *Id.* § 216(h)(4)(A), (7)(B)(ii).

face and its likelihood of eventual success.[354] If states joined in this process they could educate the developer early on about the aspects of its proposal that might create conflict with local residents. The utility could then craft its proposal to minimize opposition. Unfortunately, there is currently no indication as to whether states will be interested in joining with the federal government in this endeavor. It is unclear what effect these provisions will have on NIMBYism and state parochialism.

C. Interstate Disagreements

Conflicts inevitably arise when many state and federal agencies must independently authorize the construction of an interstate transmission line. These conflicts have resulted in substantial delays and added expenses for large interstate proposals.[355] The sheer number of decision-makers, disparate priorities, and varying standards create these disagreements; integration of the decision-making process under a central authority should resolve many of the existing problems.

The 2005 Act attempts to reduce the conflict between various authorizing agencies by combining all federal authorizations into a single review process that state and regional agencies are permitted to join.[356] This should be a very effective means of preventing disagreement between different agencies, and the states that choose to join. As there is only one authorization, by definition the parties involved must reach an agreement. With a lead agency and binding deadlines,[357] major delays should be a thing of the past. If a deadline is missed or authorization is refused, the applicant can appeal to the President, who must make a decision within 90 days.[358]

The major questions are how the provision will be implemented and whether the states will choose to participate. The implementation provisions are vague, requiring only that the Secretary promulgate whatever regulations are necessary and that involved agencies enter a memorandum of understanding. Clearly the efficacy of this new streamlined authorization will depend to a large extent on what these regulations mean. States may be willing to join this process because it could make it much easier for them to cooperate in the siting of electric transmission facilities.

If states cede their siting responsibilities to Regional Transmission Siting Agencies as envisioned by the Act, interstate friction would be substantially reduced in the permitting process as well. With siting decisions made at a regional level, states would be forced to cooperate in decision-making. If all states join these agencies and the one-stop federal authorization is well-implemented, interstate disagreement could disappear from transmission

354. *Id.* § 216(h)(4)(C).
355. *See supra* discussion Part II.A.
356. 2005 Act, *supra* note 1, § 216(h)(1)-(9).
357. *Id.* § 216(h)(2), (4)(A).
358. *Id.* § 216(h)(6)(A)-(C).

siting. If states choose not to participate in either development, state disagreements will still be a negative factor in transmission facility siting. Nevertheless, the situation should at least be improved, and assuming the proper streamlining of federal authorization and mandated regional siting, conflicts would be dramatically reduced.

V. CONCLUSION

"Modern society has come to depend on reliable electricity as an essential resource for . . . nearly all aspects of modern life."[359]

The shortage of electric transmission infrastructure in the United States is rapidly approaching crisis levels. Deregulation has created a thriving wholesale electricity market that threatens to overwhelm a transmission grid built in the days of monopoly utilities and isolated local power flows. Investment in new transmission capacity is presently deterred by archaic transmission facility siting regulations that make it difficult to site interstate transmission lines.

The industry has clamored for legislation that would transfer siting authority from state and local governments to regional or national entities that can adequately account for the vast regional benefits of interstate transmission lines. State organizations and officials, on the other hand, have protested against any such measures, proclaiming that transmission siting is an inherently local concern.

With the 2005 Energy Policy Act, Congress passed a compromise. The Act allows for federal siting authority, but only in areas where congestion on the electrical grid has become a national concern, and only in states that are not authorized to take regional benefits into account or that have withheld their approval of a proposal for more than a year. The Act also offers no specifics as to the manner in which the federal siting process will be implemented. The Act encourages states to cede their siting authority to regional agencies, but the only incentive it provides is exclusion from its federal siting provisions. With the impact and extent of federal siting authority so uncertain, only time will tell how much pressure can be brought to bear on the states to approve critical transmission proposals in a timely fashion and to cede their siting responsibilities to regional entities. Regional siting authority would be an improvement, but if the Act is well implemented it has the potential to encourage investment in new transmission capacity and to stave off a catastrophic electric transmission shortage.

359. BLACKOUT REPORT, *supra* note 67, at 5.

BRIDGING THE PHYSICAL – MENTAL GAP: AN EMPIRICAL LOOK AT THE IMPACT OF MENTAL ILLNESS STIGMA ON ADA OUTCOMES

WENDY F. HENSEL*

GREGORY TODD JONES**

INTRODUCTION

On July 26, 1990, President George H. W. Bush signed the Americans with Disabilities Act[1] ("ADA") into law, offering expansive protection from discrimination for the first time to an estimated forty-three million Americans with disabilities.[2] Members of a variety of disability-rights organizations were present at the ceremony, representing the collective victory achieved for all people with disabilities.[3] Many have credited the group consciousness in evidence that morning as a primary motivating factor in securing passage of the ADA.[4] From the time of the bill's conception, disability-rights advocates urged people with disabilities to look beyond their divergent medical conditions toward the collective stigmatization that all group members experience.[5] They clearly communicated that in the drive to secure legislation,

---

* Assistant Professor of Law, Georgia State University College of Law. The author would like to thank Mary Radford, Roy Sobelson and Kelly Timmons for their valuable advice on earlier drafts of this article.

** Director of Research, Consortium on Negotiation and Conflict Resolution, and Faculty Research Fellow, Georgia State University College of Law.

1. The Americans with Disabilities Act (ADA), 42 U.S.C. §§ 12101-213 (2000).

2. 42 U.S.C. § 12101(a)(1) (finding that "some 43,000,000 Americans have one or more physical or mental disabilities").

3. NATIONAL COUNCIL ON DISABILITY, EQUALITY OF OPPORTUNITY: MAKING OF THE AMERICANS WITH DISABILITIES ACT 178 (1997) [hereinafter EQUALITY OF OPPORTUNITY] (noting that approximately 3,000 people with and without disabilities were at the White House ceremony). In his remarks at the ADA's signing, President Bush thanked the "dedicated organizations for people with disabilities who gave their time and their strength and, perhaps most of all, everyone out there across the breadth of this nation, the 43 million Americans with disabilities. You have made this happen. All of you have made this happen." *Id.* at 231.

4. *See, e.g., id.* at 183 ("First and foremost, the ADA is a tribute to the growth and organization of the disability rights movement [which] . . . asserted itself and became a political force to be reckoned with"); *cf.* Jean Campbell, *Unintended Consequences in Public Policy: Persons with Psychiatric Disabilities and the Americans with Disabilities Act*, 22 POL'Y STUD. J. 133, 138 (1994) (noting the "united resolve by the entire disability community to continue to sponsor comprehensive coverage of disability protections").

5. EQUALITY OF OPPORTUNITY *supra* note 3, at 169-171 ("ADA advocates had made a

"'[n]o subgroup of people with any type of physical or mental disability, or perceived disability, no matter how politically impotent or how stigmatized, will be sacrificed.'"[6] Because these disparate constituencies successfully focused on a shared history and contemporaneous experience of external discrimination, they were able to effectively organize a collective body insistent on rights for the whole.

Nevertheless, the disparate nature of group members' interests can be both problematic and unclear. People with disabilities diverge from each other in material ways that affect both their legal interests and access to political power.[7] There is a hierarchy of position and social acceptance within the disabled community, and in virtually every meaningful way, individuals with psychiatric disorders[8] fall at the bottom of the pecking order.[9] The stigma that continues to adhere to mental illness impedes the recognition of people with disabilities as a collective social minority and encourages individuals with physical impairments to distance themselves from their mentally impaired

---

commitment more than a year before that they would stand together: one for all and all for one.").

6. *From ADA to Empowerment: The Report of the Task Force on the Rights and Empowerment of Americans with Disabilities* 14 (October 12, 1990), *quoted in* EQUALITY OF OPPORTUNITY, *supra* note 3, at 81. This unity was particularly evident during debates over the proposed Chapman Amendment to the ADA, which would have permitted restaurants to remove HIV positive food handlers from their positions. EQUALITY OF OPPORTUNITY, *supra* note 3, at 169-171.

7. *See* Campbell, *supra* note 4, at 134 ("[D]isabled persons are not a homogeneous group with common needs and equal social power . . . .").

8. The terms "psychiatric disorder" and "mental illness" are used synonymously throughout this Article. A broad range of mental and emotional conditions that affect behavior, mood, and cognition are intended by these general terms, including depression, anxiety disorders, and schizophrenia. A distinguishing characteristic of such disorders is that they are predominantly treated through the use of psychotropic medication. *See* Karin A. Guiduli, *Challenges for the Mentally Ill: The "Threat to Safety" Defense Standard and the Use of Psychotropic Medication under Title I of The Americans with Disabilities Act of 1990*, 144 U. PA. L. REV. 1149, 1151 n.9 (1996) (citing definitions in DEBORAH ZUCKERMAN ET AL., THE ADA AND PEOPLE WITH MENTAL ILLNESS: A RESOURCE MANUAL FOR EMPLOYERS 1 (1993)). Significantly, these terms for purposes of this Article do not include disorders arising from organic brain damage, mental retardation, or similar neurological conditions, which nevertheless would fall within the broader definition of mental impairment under the ADA. Although some studies have treated alcoholics and substance abusers as mentally ill, they generally are not included within the scope of this Article except where specifically identified.

9. Michael L. Perlin, *The ADA and Persons with Mental Disabilities: Can Sanist Attitudes Be Undone?*, 8 J. L. & HEALTH 15, 20 (1994) ("Even within the disability community, persons with mental illness are often the poor stepchild, and remain the last hidden minority."); Carol L. Owen, To Tell or Not to Tell: Disclosure of a Psychiatric Condition in the Workplace 42 (2004) (unpublished Ph.D. dissertation, Boston University School of Arts and Sciences) (on file with UMI Dissertation Services) ("[P]eople with psychiatric disabilities . . . [are] the most significantly stigmatized group among people with disabilities . . . .").

brethren.[10]

Some scholars have argued that this prejudice is reflected in both the legislative history and subsequent application of the ADA.[11] Unlike other civil rights laws, each individual seeking protection under the ADA must first make a threshold showing of membership in the protected class.[12] The court will not consider the defendant's allegedly discriminatory actions unless and until the plaintiff first establishes that he or she is a qualified individual with a disability within the meaning of the Act.[13] There is no question that all plaintiffs have experienced difficulty in making this showing as a result of several Supreme Court decisions narrowing the definition of "disability."[14] Many scholars have argued that courts do not appreciate the social construction of disability and focus too much attention on medical diagnoses and functional limitations rather than the impact of societal attitudes and prejudices.[15] Some scholars have gone further to theorize that individuals alleging mental impairments face even greater challenges than physically-impaired litigants in ADA cases, in large part due to the persistent stigma attached to group members by many Americans, including judges.[16]

---

10. *See, e.g.*, Jane Byeff Korn, *Crazy (Mental Illness Under the ADA)*, 36 U. MICH. J.L. REFORM 585, 601-02 (2003) (noting that people with mental disabilities "feel that they are discriminated against even within the disability community").

11. *See, e.g., id.* at 652 ("[T]he ADA is being used to continue the current second-class status of mental illness."); Campbell, *supra* note 4, at 134, 137 (arguing that "the subtext of the ADA legislation reveals stereotypes and discrimination" and noting that literature on accommodation focuses on physical, not psychiatric disabilities); Perlin, *supra* note 9, at 19-22 (questioning the ADA's impact on people with mental illness).

12. *See, e.g.*, Robert L. Burgdorf, Jr., *The Americans with Disabilities Act: Analysis and Implications of a Second-Generation Civil Rights Statute*, 26 HARV. C.R.-C.L. L. REV. 413, 441-42 (1991) (comparing the class-based structure of the ADA to other civil rights statutes).

13. 42 U.S.C. § 12112 (2000).

14. *See, e.g.*, Toyota Motor Mfg., Ky., Inc. v. Williams, 534 U.S. 184, 200-02 (2002) (finding that an employee who was unable to perform repetitive manual tasks on the job was not disabled in part because she was able to perform household tasks); Sutton v. United Air Lines, Inc., 527 U.S. 471, 482 (1999) (holding that courts must take into account all mitigating measures employed by plaintiffs in determining whether someone is disabled within the meaning of the ADA); Murphy v. United Parcel Serv., Inc., 527 U.S. 516, 521 (1999) (relying on *Sutton*). For a general discussion of how these cases have restricted the definition of disability under the statute, see Lisa Eichhorn, *The* Chevron *Two-Step and the* Toyota *Sidestep: Dancing Around the EEOC's "Disability" Regulations Under the ADA*, 39 WAKE FOREST L. REV. 177 (2004) and Bonnie Poitras Tucker, *The Supreme Court's Definition of Disability Under the ADA: A Return to the Dark Ages*, 52 ALA. L. REV. 321 (2000).

15. *See, e.g.*, Laura L. Rovner, *Disability, Equality, and Identity*, 55 ALA. L. REV. 1043, 1044-45 (2004) ("[T]he success of the disability community in infusing the socio-political model of disability into federal law has begun to be eroded by judicial decisions interpreting the ADA that appear to be grounded in – and espousing – the medical model of disability.").

16. *See, e.g.*, Wendy F. Hensel, *Interacting With Others: A Major Life Activity Under The Americans with Disabilities Act?*, 2002 WIS. L. REV. 1139, 1168-75 (concluding that

This Article represents one of the first attempts to ascertain empirically whether mentally-impaired litigants experiencing psychiatric disorders are disadvantaged vis-à-vis those with physical impairments in establishing membership in the protected class under the ADA. Part I provides the background for the study by identifying and discussing the historical stigma and stereotypes associated with mental illness and psychiatric impairments which continue to thrive today. Particular attention is paid to the potential impact such stigma may have in an employment context, and the challenges it creates for impaired employees in the workplace. Part II scrutinizes the legislative history of the ADA to determine whether evidence of such stigma was present during the legislative process, identifying the reluctance of several lawmakers to extend protection to people with psychiatric impairments. Part III reveals the results of an empirical study of all Title I district and circuit court decisions, available in LexisNexis or Westlaw, in three appellate circuits over two discrete two-year time periods, which directly address whether litigants have established class coverage under the ADA. The study compares the success rates of physically and mentally impaired litigants in an attempt to answer whether individuals with psychiatric disorders face systemic challenges that are not shared by physically-impaired litigants. Part IV proposes possible explanations for the surprisingly comparable success rates between the two tested groups, examining subtle sources of continuing stigma that may explain the unexpected results. Part V concludes with proposals for future research and action by advocates to ensure that people with mental illness are able to secure the full and equal protection from discrimination that is the promise of the ADA.

I. STIGMA AND STEREOTYPES: PERCEPTIONS OF INDIVIDUALS WITH MENTAL ILLNESS

It is no secret that individuals labeled mentally ill by society have been stigmatized and subjected to discriminatory treatment throughout history.

---

judicial bias may play a role in the higher success rates experienced by litigants with physical impairments in establishing that interacting with others is a major life activity); Korn, *supra* note 10, at 588-89, 642 (stating that people with mental disabilities are "in an even worse position" in establishing a protected disability than those with physical disabilities, "due, in large part, to the stereotypes surrounding mental illness"); Susan Stefan, *"You'd Have to be Crazy to Work Here": Worker Stress, the Abusive Workplace, and Title I of the ADA*, 31 LOY. L.A. L. REV. 795, 805 (1998) (contending that "judicial assumptions about the nature of psychiatric disabilities and essential employment functions have resulted in the near-total failure of the ADA to protect individuals with psychiatric disabilities from employment discrimination"); Michelle Parikh, Note, *Burning the Candle at Both Ends, and There is Nothing Left for Proof: The Americans with Disabilities Act's Disservice to Persons with Mental Illness*, 89 CORNELL L. REV. 721, 759 (2004) ("Courts show a great discomfort in dealing with these [mental illness] cases, and apparently intuit that mentally ill individuals should not recover for employment discrimination.").

Although individuals with physical impairments have also been the subject of disparaging public opinion, the animus directed at psychiatric impairments is proportionately greater and more pervasive.[17] Individuals with psychiatric disorders repeatedly experience restricted access to social interactions with others and to fundamental goods and services, like employment and housing.[18] Society's growing acceptance of psychology and the law's current protection of at least some group members have not greatly altered this unpromising landscape.[19] One study recently concluded that individuals today may be more likely to hold stigmatizing attitudes toward the mentally ill than their counterparts nearly five decades ago.[20] The unemployment rate of individuals with mental illness continues to be "three to five times higher than among those with no psychiatric disorder,"[21] reportedly the "'worst level of employment of any group of people with disabilities.'"[22] Repeated studies have also shown that landlords are less likely to lease to this group, making it difficult

---

17. *See* Patrick W. Corrigan & David L. Penn, *Lessons from Social Psychology on Discrediting Psychiatric Stigma*, 54 AM. PSYCHOLOGIST 765, 766 (1999) ("[T]he general public seems to disapprove of persons with severe mental illness significantly more than of persons with physical disabilities like Alzheimer's disease, blindness, or paraplegia."); Korn, *supra* note 10, at 586-87 ("Whatever stigma attaches to a physical disability, however, most scholars agree that people with mental disabilities are more feared, more stigmatized, discriminated against more often, and are seen as more likely to commit acts of violence than are people with physical disabilities.").

18. *See* SUSAN STEFAN, UNEQUAL RIGHTS: DISCRIMINATION AGAINST PEOPLE WITH MENTAL DISABILITIES AND THE AMERICANS WITH DISABILITIES ACT 4-25 (2001) (detailing the pervasive effects of discrimination against people with psychiatric disabilities); Patrick W. Corrigan & Petra Kleinlein, *The Impact of Mental Illness Stigma, in* ON THE STIGMA OF MENTAL ILLNESS: PRACTICAL STRATEGIES FOR RESEARCH AND SOCIAL CHANGE 11, 18-20 (Patrick W. Corrigan ed., 2005) (detailing challenges faced by individuals labeled as mentally ill).

19. Jack K. Martin et al., *Of Fear and Loathing: The Role of 'Disturbing Behavior,' Labels, and Causal Attributions in Shaping Public Attitudes Toward People with Mental Illness*, 41 J. HEALTH & SOC. BEHAV. 208, 219 (2000) (noting that the results of the authors' study "do not support the contention that the consequences associated with mental health problems have dissipated.").

20. Jo C. Phelan et al., *Public Conceptions of Mental Illness in 1950 and 1996: What is Mental Illness and is it to be Feared?*, 41 J. HEALTH & SOC. BEHAV. 188, 197-98 (June 2000).

21. Corrigan & Kleinlein, *supra* note 18, at 18 (citing study by Roland Sturm et al., *Labor Force Participation by Persons with Mental Illness*, 50 PSYCHIATRIC SERVICES 1407 (1999)). The unemployment rate for people with severe mental illness is generally estimated at 85%. *Id.*; *see also* SUSAN STEFAN, HOLLOW PROMISES: EMPLOYMENT DISCRIMINATION AGAINST PEOPLE WITH MENTAL DISABILITIES 12 (2002).

22. Patrick J. Kennedy, *Why We Must End Insurance Discrimination Against Mental Health Care*, 41 HARV. J. ON LEGIS. 363, 372 (2004) (quoting PRESIDENTS'S NEW FREEDOM COMM'N ON MENTAL HEALTH, INTERIM REPORT TO THE PRESIDENT 11, (Oct. 29, 2002)). The President appointed the Commission "to examine barriers to mental health care and to make recommendations for improvements." *Id.*

to secure adequate housing.[23] Significantly, these effects exist whether or not the individual labeled as mentally ill has exhibited psychiatric symptoms or engaged in stigmatizing behavior.[24]

In the following subsections, this Article explores the most common negative stereotypes that society has attached to individuals with mental illness. This contextual history sheds light on the challenges faced by individuals seeking protection for psychiatric disorders under the ADA.

A. Dangerousness and Unpredictability

Perhaps no stereotype is more embedded in the American psyche than that of the dangerously unpredictable mental patient.[25] Concerns with dangerousness vary to some extent depending on the type of psychiatric disorder alleged.[26] Studies have demonstrated, however, that society's overall concern with the potential for violence is high regardless of the mental impairment in question,[27] and actually has increased over the last 40 years.[28] Such concerns are both mirrored in and fueled by news coverage of mental health matters, which is consistently dominated by discussions of violence and crime committed by the mentally ill.[29] The vast majority of entertainment productions featuring the mentally ill reflect these results. One study found

---

23. *See* Patrick W. Corrigan et al., *From Whence Comes Mental Illness Stigma?*, 49 INT'L J. OF SOC. PSYCHIATRY 142, 144 (2003) (citing studies to this effect).

24. *Id.* ("[A] growing body of experiential and empirical data [exists] that impl[ies] people with mental illness will experience discrimination regardless of their behavior.").

25. *Cf.* Bernice A. Pescosolido et al., *The Public's View of the Competence, Dangerousness, and Need for Legal Coercion of Persons with Mental Health Problems*, 89 AM. J. PUB. HEALTH 1339, 1339 (1999) (explaining that one of the three core assumptions underlying "[m]odern public health laws dealing with mental disorders is that mental disorders may place a person at increased risk of physically harming himself or herself or others").

26. Martin et al., *supra* note 19, at 215. The study found that society is most afraid of individuals with substance abuse problems. *See id.*

27. Pescosolido, *supra* note 25, at 1341. The study asked a sample of individuals to consider a variety of vignettes involving individuals with mental health problems and answer questions about how likely it was that the individual would "do something violent toward others." *Id.* The authors reported that "[a]lmost 17% of the sample indicated that even the 'troubled person' was either very or somewhat likely to do something violent toward others. That percentage rose to 33.3% for the depression vignette and to more than 60% for the schizophrenia vignette." *Id.*; *see also* Martin et al., *supra* note 19, at 220 (concluding that data gathered in the authors' study reflected "a continuing widespread concern that people with mental health problems are likely to be dangerous to others").

28. Phelan, *supra* note 20, at 197-98.

29. *See* Otto F. Wahl, *News Media Portrayal of Mental Illness: Implications for Public Policy*, 46 AM. BEHAV. SCIENTIST 1594, 1595-96 (2003) (detailing the results of several media studies). This is true despite the fact that a relatively small percentage of crime is committed by the mentally ill. *See* Korn, *supra* note 10, at 586 ("[M]ore than 95% of violent acts in our society are committed by people who are *not* mentally ill.").

that 72.1% of all characters in primetime television who were portrayed as mentally ill either hurt or killed others.[30] Children's programming is no exception, with one study finding that Walt Disney animated films often portray the mentally ill as "objects of . . . fear."[31] This cultural vision persists despite several studies indicating that the association between mental disorders and the risk of violence is, at best, "weak" or "modest."[32]

Perceptions of dangerousness lead many to fear those with psychiatric disorders, which in turn drives their desire to increase social distance from group members.[33] Such fears can be a significant impediment to those seeking to enter the workforce or to maintain employment once secured.[34] Almost 60% of people surveyed in a recent study expressed their unwillingness to work alongside someone who suffers from a substance abuse problem, schizophrenia, or depression.[35] Employers reflect the same uncertainties and are reluctant to employ any individual perceived as posing a threat to co-workers.[36] Aside from concerns about violence in the workplace, employers worry about the potential for open-ended liability if an incident does occur, and the charge that they negligently failed to stop what is perceived as "preventable and predictable" violence.[37] Even an individual's solid work history and impressive credentials are unlikely to outweigh these concerns for most employers.[38] Despite efforts by public interest campaigns and mental health advocates to combat perceptions of dangerousness, these images remain deeply entrenched, even among many mental health professionals.[39]

---

30. Donald L. Diefenbach, *The Portrayal of Mental Illness on Prime-Time Television*, 25 J. COMMUNITY PSYCHOL. 289, 290 (1997), *citing* Nancy Signorielli, *The Stigma of Mental Illness on Television*, 33 J. BROADCASTING & ELECTRONIC MEDIA 325, 327-28 (1989).

31. *See* Andrea Lawson & Gregory Fouts, *Mental Illness in Disney Animated Films*, 49 CAN. J. PSYCHIATRY 310, 313 (2004).

32. *See* Ann Hubbard, *The ADA, the Workplace, and the Myth of the "Dangerous Mentally Ill,"* 34 U.C. DAVIS L. REV. 849, 852, 867-73 (2001) (discussing several studies on the relationship between violence and mental illness).

33. Phelan, *supra* note 20, at 202 (citing a study by Bruce G. Link et al., *The Social Rejection of Former Mental Patients: Understanding Why Labels Matter*, 92 AM. J. SOC. 1461 (1987)).

34. *See* Martin et al., *supra* note 19, at 216.

35. *Id.*

36. *See* Wahl, *supra* note 29, at 1596.

37. Korn, *supra* note 10, at 613-14.

38. *See generally* Wahl, *supra* note 29, at 1596.

39. *See* Corrigan & Penn, *supra* note 17, at 766 (describing studies finding that even "well-trained professionals from most mental health disciplines subscribe to stereotypes about mental illness").

B. Incompetence

Another pervasive negative stereotype associated with mental illness is that of the incompetent who engages in childlike behavior.[40] Others believe that they may justifiably make life decisions for individuals with psychiatric disorders without taking their wishes into consideration because group members are irresponsible and unable to care adequately for themselves.[41] Although this model does not inspire the same need for social distance as the images of dangerousness discussed above, its paternalistic benevolence is equally demoralizing to group members who can never aspire to the role of equal or peer with others, let alone expect to be treated as a friend.[42]

This imagery is understandably problematic in a workplace setting, where hiring and advancement depend upon an employer's confidence in the abilities of an employee. Such confidence may be radically undermined or eliminated altogether once an employee is labeled mentally ill, either through an official diagnosis or through the office grapevine.[43] One recent study supports this conclusion, finding that over 70% of employees believed that their employers treated them as "less competent" following the disclosure of a mental impairment in the workplace.[44] Far from protecting group members, such imagery is seriously detrimental to their ability to function adequately in society.

C. Defective Character

Throughout history, society has viewed mental illness as internally generated and the product of poor character.[45] This negative imagery does not extend to people with physical impairments, who are largely viewed as possessing involuntary and uncontrollable disorders.[46] This internal attribution

---

40. *See id.*; Pescosolido et al., *supra* note 25, at 1339 (concluding that regardless of the severity of the identified disorder, the public views individuals with a mental illness as less competent).

41. Corrigan and Penn, *supra* note 17, at 766.

42. *Id.* at 766-67.

43. *See* Bruce Link, *Mental Patient Status, Work, and Income: An Examination of the Effects of a Psychiatric Label*, 47 AM. SOC. REV. 202, 210 (1982) ("Other things equal, including psychiatric condition, the acquisition of the status of mental patient appears to have a negative impact on a person's chances of getting and keeping a job.").

44. Otto Wahl, MEDIA MADNESS: PUBLIC IMAGES OF MENTAL ILLNESS (Rutgers University Press 1995), *cited in* Martin et al., *supra* note 19, at 210.

45. Historically, individuals with mental illness were viewed as being "demon-possessed" and without "sufficient moral backbone to hold off Lucifer." Patrick W. Corrigan & Alicia K. Matthews, *Stigma and Disclosure: Implications for Coming Out of the Closet*, 12 J. MENTAL HEALTH 235, 238 (2003).

46. *See* Corrigan & Penn, *supra* note 17, at 766.

of blame can be devastating because the level of stigma that society attaches to a group is directly linked to society's beliefs about the causes of the undesired behavior or traits associated with group members.[47] Society is most tolerant of negative characteristics caused by factors beyond an individual's control, such as genetics, educational disadvantage, or the environment.[48] On the other hand, characteristics perceived to be within the individual's control are equivalent to character flaws that intensify stigma and the desire for social distance.[49] Because the characteristic is controlled by the individual, society views the failure to conform to its standards as the product of unwillingness, rather than inability.[50] If the individual would only choose to try harder, all problems would disappear.[51] Rather than viewing group members with pity, an emotion often associated with physical disabilities, society instead views them with anger and irritation.[52] The desire to offer assistance in any respect is seriously diminished because the need only arises as a result of the individual's failure to help himself.[53]

Some evidence suggests that more Americans today believe there are biological causes underlying mental illness, a view supported in part by the increasingly public disclosures of mental illness by prominent Americans and public awareness campaigns.[54] This perspective, however, is not universal. [55] As one poll reported, almost 50% "of Americans think that depression is caused by the weak character of the sufferer" rather than by any external factors.[56] Unless and until such perceptions change, individuals with psychiatric disorders will be deemed the cause of their own problems.

D. Malingerers

Perhaps the flip side to the image of the mentally ill as poor in character is the image of the active deceiver, who in an employment context creates or exaggerates the effect of a mental disorder to secure favorable workplace

---

47. Martin et al., *supra* note 19, at 211, 218.

48. *Id.* at 218.

49. *Id. Cf.* Ladonna L. Rush, *Affective Reactions to Multiple Social Stigmas*, 138 J. SOC. PSYCHOL. 421, 427 (1998) (finding in a controlled study that participants' views of various stigmatized groups were influenced by the perceived controllability of their stigmatized characteristics).

50. Stefan, *supra* note 18, at 9-10.

51. *Id.* at 10.

52. *See id.* at 9; Corrigan & Penn, *supra* note 17, at 766.

53. Patrick W. Corrigan et al., *Challenging Two Mental Illness Stigmas: Personal Responsibility and Dangerousness*, 28 SCHIZOPHRENIA BULL. 293,293-94 (2002).

54. *See, e.g.*, Julie Solomon & Claudia Kalb, *Breaking the Silence*, NEWSWEEK, May 20, 1996, at 20 (discussing revelations by politicians, actors, and other high profile Americans about mental illness that either they or someone in their families have experienced).

55. Martin et al., *supra* note 19, at 216-17.

56. Stefan, *supra* note 18, at 9.

accommodations or to avoid discipline. Mental impairments are neither easily identified by the average person nor, in most cases, static and permanent. Society, however, "is most comfortable with disabilities that are permanent and chronic; either one is disabled or is not."[57] This is reflected in the international symbol of disability – an individual in a wheelchair. The untrained layperson feels confident that he understands the nature of such limitations, which appear both real and obvious in his eyes. In contrast, because mental illness is largely diagnosed on the basis of patients' subjective reports and doctors' observations, society questions the legitimacy of such claims, especially when the individual stands to benefit.[58]

Compounding the problem of intangibility is the reality that most people do not experience mental illness as a constant, but instead as episodes of crises alternating with long periods of productive functioning and achievement.[59] These periods of well-being are not viewed as evidence of progress, however, but as concrete proof that no disability existed in the first place. Employees seeking accommodation are viewed with heavy skepticism and may be regarded by employers as malingerers seeking to avoid work or the consequences of their own actions.[60] The widespread nature of this caricature is confirmed by its appearance both in the popular press[61] and in works by law professors.[62]

II. LEGISLATIVE HISTORY OF THE ADA

The pervasiveness of the stigma attached to mental illness is reflected in the legislative debates surrounding the passage of the ADA. Although much

---

57. *Id.* at 10.

58. *See* Korn, *supra* note 10, at 595.

59. Stefan, *supra* note 18, at 58.

60. Teresa L. Scheid, *Employment of Individuals with Mental Disabilities: Business Response to the ADA's Challenge*, 17 BEHAV. SCI. L. 73, 75 (1999).

61. Wahl, *supra* note 29, at 1596. Wahl describes a newspaper cartoon that appeared shortly after the EEOC issued guidelines about the ADA's application to people with mental disabilities as follows:

> The cartoon showed a man with a briefcase labeled EEOC speaking to an employer. He is telling the employer, "No you may not ask a job applicant about a history of mental disabilities. That's discrimination." The prospective employee standing next to the EEOC representative is wearing a hockey mask and holding a raised axe.

Id.; *see also* Richard L. Lesher, *Disabilities Act is a Failed Experiment*, 51 HUMAN EVENTS 17-18 (1995) (noting that very few cases brought under the ADA "involve plaintiffs afflicted with legitimate *disabilities* as most of us understand [them]," and as a result, "it is clear the ADA has become yet another weapon in the hands of disgruntled employees with personal axes to grind and little inclination to accept personal responsibility for their own condition").

62. *See, e.g.*, Stefan, *supra* note 18, at 12 (discussing Alan Dershowitz, THE ABUSE EXCUSE AND OTHER COP-OUTS, SOB STORIES, AND EVASIONS OF RESPONSIBILITIES (Little Brown 1994)).

has been written about the ADA's journey through Congress from concept to reality, less attention has focused on the reluctance expressed by some members of Congress to provide employment protection to individuals suffering from psychiatric disorders. Because courts use legislative history to guide and inform their interpretations of statutory law, understanding these concerns may shed light on the subsequent approach taken by judges in interpreting the meaning of disability and the prima facie case under the ADA.

The first draft of the ADA was introduced to the 100th Congress in late April 1988, when few expected it to pass.[63] Ironically, President Reagan inadvertently fueled the growing demand for disability discrimination legislation in the early stages of the debate by characterizing the Democratic nominee for President, Michael Dukakis, as mentally ill.[64] When asked to comment on rumors that Dukakis had "sought psychological help [for depression]," Reagan smiled and stated that he was "not going to pick on an invalid."[65] His remark was immediately criticized by the media as well as members of Congress, who denounced "the callous attitude exhibited by the Reagan-Bush administration toward those with disabilities, of which this remark is symptomatic."[66] Despite protests over the unfairness and inaccuracy of Reagan's comments, Dukakis quickly dropped in the polls, reflecting the difficulties often experienced by political candidates tied to mental illness, real or otherwise.[67] Shortly after the controversy made headlines, however, then-candidate George H. W. Bush announced his support for the ADA, and once elected, directed his staff to work with Congress to aid its enactment.[68] At

---

63. See H.R. 4498, 100th Cong. (1988). Senator Lowell Weicker (R-CT) introduced the first version of the ADA (S. 2345) on April 28, 1988. EQUALITY OF OPPORTUNITY, *supra* note 3, at 207.

64. Jacob V. Lamar, *Reagan: Part Fixer, Part Hatchet Man*, TIME MAGAZINE, Aug. 15, 1988, at 16.

65. *Id.*

66. *Id.*; 134 CONG. REC. 21,425 (1988) (statement of Rep. Coelho (D-CA)).

67. One study conducted by the Depression and Bipolar Support Alliance found that, in a race for national office where all other factors were equal, nearly one in four (24%) would vote against a candidate who had been diagnosed with clinical depression, while another 24% said they "might not" vote for them. It also found that nearly one in three (31%) believe that people with mood disorders are not stable enough to hold positions of authority in fields like law enforcement and government, while half (51%) feel that people should publicly disclose such diagnoses if they seek office.

Depression and Bipolar Support Alliance, *Thirty Years After Eagleton Controversy, Mental Health Stigma Still Haunts Political Arena*, http://www.dbsalliance.org/media/newsreleases/Stigma%20Release.html (last visited Feb. 3 , 2006).

68. *See* RUTH COLKER, THE DISABILITY PENDULUM: THE FIRST DECADE OF THE AMERICANS WITH DISABILITIES ACT 31, 67-68 (2005) (noting that the ADA gained momentum from "President Reagan's insensitive comments" because they "helped spur presidential candidate Bush to endorse the concept of the ADA, and later, to instruct his attorney general to seek its passage in Congress").

least one scholar has credited the pressure generated from Reagan's comments as facilitating Bush's political support of the statute.[69]

The bill was re-introduced to the 101st Congress in a modified form after consideration by numerous House and Senate committees.[70] The Senate floor debate reflected the concern of several members that the legislative definition of disability was "extremely loose" and covered too many people.[71] Although Senators most often objected to the Act's real or imagined coverage of alcoholics, drug users, and homosexuals, some members specifically identified concerns with including the mentally ill within the Act's protection.[72] A common theme in such comments was the distinction between physical and mental impairments, and the morally problematic character of the latter.

Senator Jesse Helms (R-NC) acknowledged that the ADA's protection of "people in a wheelchair or those who have been injured in the war . . . is one thing," but questioned whether it also covered manic depressives and schizophrenics.[73] He indicated that "an employer's own moral standards [should] enable him to make a judgment about any or all of the [potentially protected] employees."[74] Senator Rudy Boschwitz (R-MN) likewise sought assurances from the bill's sponsors that in cases involving applicants with "mental handicaps" and "infirmit[ies]," employers would still be permitted to make "subjective judgments" about whether the individual had the ability to do the work.[75] Although Senator Tom Harkin (D-IA) reminded his colleagues that Congress needed to extend protection beyond those in wheelchairs because "[p]eople can be mentally handicapped as well,"[76] Senator Gordon Humphrey (R-NH) later reiterated Helms' concern about the bill's coverage.[77] He noted that "we are not simply talking about the blind, the deaf, or persons confined to wheelchairs," but a bill "so broad" that it covers "schizophrenics,

---

69. *Id.*

70. Senator Tom Harkin (D-IA) and Representative Tony Coelho (D-CA) jointly introduced the revised version of the bill (S. 933 and H.R. 2273) on May 9, 1989. EQUALITY OF OPPORTUNITY, *supra* note 3, at 207.

71. 135 CONG. REC. 19,840-41 (1989) (statement of Sen. Pryor (R-AK)) (arguing that the definition of disability, potentially a "lawyer's mecca," would "be the subject of literally countless issues of litigation in the courts across this country").

72. *See, e.g., id.* at 19,864 (statement of Sen. Helms).

73. *Id.* at 19,864, 19,866. (statement of Sen. Helms). In addition to objecting to including certain mental impairments under the ADA, Senator Helms was particularly virulent in insisting that the ADA exclude homosexuals and those infected with HIV from its protection. *Id.* at 19,866 (statement of Sen. Helms).

74. *Id.* at 19,864 (statement of Sen. Helms).

75. *Id.* at 19,850 (statement of Sen. Boschwitz). Interestingly enough, Senator Edward Kennedy (D-MA) immediately confirmed that Senator Boschwitz's understanding was "correct." *Id.*

76. *Id.* at 19,866 (statement of Sen. Harkin).

77. *Id.* at 19,882 (statement of Sen. Humphrey).

manic depressives, and persons with extremely low IQ's."[78]

Senator William Armstrong (R-CO) was perhaps the most specific in his concerns about the coverage of mental impairments. He agreed that the bill was welcome to the extent it would "help[] . . . people in wheelchairs or who have some kind of a physical disability or handicap of some sort," but was concerned that the term "disability might include some things which by any ordinary definition we would not expect to be included," like "[m]ental disorders, such as alcohol withdrawal, delirium, hallucinosis, dementia with alcoholism, marijuana, delusional disorder, cocaine intoxication, cocaine delirium, [and] disillusional disorder."[79] He was particularly troubled that the ADA would provide protection to someone with disorders that "have a moral content to them or which in the opinion of some people have a moral content."[80] He sought specific assurances that a host of psychiatric disorders were not covered under the bill, including "impulse control disorders" and "disruptive behavior disorder," and offered an amendment to clarify the record if there was "any doubt" they would be covered.[81] Arguing that the bill was simply "far too broad," he also used the *Diagnostic and Statistical Manual of Mental Disorders* published by the American Psychiatric Association to demonstrate that the definition would include voyeurism if the Senators did not specifically "take it out."[82]

Even one of the ADA's co-sponsors expressed reservations about covering mental impairments.[83] After announcing on the Senate floor that he was pleased to cosponsor the bill, Senator Warren Rudman (R-NH) stated his concerns with coverage as follows:

> [W]hile our knowledge of psychiatry has greatly improved in recent years, the fact remains that a diagnosis of certain types of mental illness is frequently made on the basis of a pattern of socially unacceptable behavior and lacks any physiological basis. In short, we are talking about behavior that is immoral, improper, or illegal and which individuals are engaging in of their own volition, admittedly for reasons we do not fully understand. . . .
>
> In principle, I agree with the concept that the mentally ill should be protected from infidious [sic] discrimination just as the physically handicapped should be. However, people must bear some responsibility for the consequences of their own actions.[84]

---

78. *Id.*

79. *Id.* at 19,852-53 (statement of Sen. Armstrong).

80. *Id.* at 19,853 (statement of Sen. Armstrong).

81. *Id.* Senator Armstrong was concerned that the bill covered the following disorders: "homosexuality and bisexuality . . . exhibitionism, pedophilia, voyeurism, . . . compulsive kleptomania or other impulse control disorders . . . conduct disorder, [and] any other disruptive behavior disorder." *Id.*

82. *Id.* at 19,871 (statement of Sen Armstrong).

83. *Id.* at 19,895-96 (statement of Sen. Rudman).

84. *Id.*

In contrast to the discussion in the Senate, there was little floor debate in the House of Representatives about the coverage of mental impairments under the ADA. One of the only comments in this regard was made by Representative Robert Walker (R-PA), who objected to the House's failure to consider an amendment to "make certain that the psychological disorder section of the bill did not include the necessity to hire psychopaths in police departments."[85] Mirroring the Senate's concern with its potential protection of homosexuals and drug addicts, the majority of comments related to the expansiveness of the definition of disability.[86]

The lack of floor debate, however, did not necessarily reflect the House's endorsement of the inclusion of mental impairments under the ADA. The correspondence of at least one Representative during the relevant time period indicated that the ADA's protection of the mentally ill was misplaced because of the danger such individuals posed to the public. Congressman Chuck Douglas (R-NH) distributed a memo to House members with a picture of a man pointing a gun at the reader.[87] Under the caption "Berserkers: Time Bombs in the Workplace," Douglas asked: "How can we protect ourselves from an apparently growing menace?"[88] His solution was to defeat "the Americans with Disabilities Act in its present form" because it "protects people who do not have physical disabilities."[89] Because of his belief that people with psychiatric disorders could be violent towards others, Douglas advocated that the ADA exclude the mentally ill from its coverage.[90]

In the midst of such negative imagery, however, one Senator stepped forward to directly advocate for the inclusion of individuals with psychiatric disorders within the protection of the ADA. Senator Pete Domenici (R-NM) spoke eloquently about the need to eliminate the stigma of mental illness, noting that both Winston Churchill and Abraham Lincoln may have experienced manic-depressive episodes.[91] He argued that "the time has come when [those with mental impairments] deserve an unbiased evaluation of their capability" rather than exclusion from the workforce based on subjective bias.[92] Ultimately, Senator Domenici's position won the day. The images of dangerousness and moral defect resurrected during the debates were not

---

85. 136 Cong. Rec. 10,841 (statement of Rep. Walker). Representative Charles Douglas (R-NH) had proposed the amendment. *Id.*

86. *See, e.g., id.* at 11,457 (statement of Ronald Lindsey, introduced into the Congressional Record by Representative Tom Delay (R-TX), that "[t]here are over a thousand different impairments that could render a person disabled, and, therefore, protected for purposes of the ADA.").

87. EQUALITY OF OPPORTUNITY, *supra* note 3, at 133.

88. *Id.*; Campbell, *supra* note 4, at 137.

89. Campbell, *supra* note 4, at 137.

90. EQUALITY OF OPPORTUNITY, *supra* note 3, at 133.

91. 135 CONG. REC. 19,878-79 (1989) (statement of Sen. Domenici).

92. *Id.*

sufficient to sway supporters of an inclusive definition of disability. The final version of the ADA, signed into law in July 1990, included individuals with mental impairments within the protected class.[93]

III. THE AMERICANS WITH DISABILITIES ACT: AN EMPIRICAL LOOK AT THE SUCCESS RATES OF INDIVIDUALS WITH MENTAL IMPAIRMENTS

The question of who is disabled within meaning of the ADA was at the center of the legislative debate. This question remains critical because the ADA, unlike other civil rights legislation, requires all litigants seeking its protection to establish coverage under the Act as a threshold requirement.[94] The Act's prohibition on discrimination and mandate of reasonable accommodation are limited to "qualified individual[s] with a disability."[95] Identifying which individuals fall within the scope of this definition has caused much confusion and debate in both courts and academia. Some scholars argue that the stigma which continues to attach to mental illness, reflected both in studies and in the legislative debate, has led courts to scrutinize claims brought by mentally-impaired individuals more closely than those brought by their physically-impaired counterparts.[96] They argue that because litigants with mental impairments have more difficulty establishing class membership, less attention is paid to defendants' potentially discriminatory actions in these cases.[97]

The following subsections reveal the surprising and counter-intuitive results of one of the first empirical tests evaluating this theory. Details of these findings follow a brief discussion of the legal requirements necessary to establish coverage in the class protected by the ADA.

A. Establishing Coverage under the ADA

In order to state a claim under the ADA, litigants must first demonstrate that they have an actual disability within the meaning of the statute, "a record of such an impairment," or that they are "regarded as having such an impairment."[98] All three approaches require evidence of a physical or mental

---

93. 42 U.S.C. § 12102 (2000) (including "physical or mental impairment" in the definition of disability); 42 U.S.C. § 12112 (2000)("No covered entity shall discriminate against a qualified individual with a disability because of the disability of such individual . . . .").

94. 42 U.S.C. § 12112.

95. 42 U.S.C. § 12112(a).

96. *See, e.g.*, Korn, *supra* note 10, at 641-42.

97. *Cf.* Jane Bryant Korn, *Cancer and the ADA: Rethinking Disability*, 74 S. CAL. L. REV. 399, 416 (2001) ("If a plaintiff cannot prove membership in the protected class, the case is dismissed; the actions of the defendant are irrelevant.").

98. 42 U.S.C. § 12102(2)(A)-(C) (2000); 29 C.F.R. § 1630.2(g)(1)-(3) (2004). It is worth noting that the Supreme Court has questioned the EEOC's authority to interpret the meaning of "disability" because it is defined in the general provisions of the ADA rather than in Title I.

impairment "that substantially limits . . . [a] major life activit[y]."[99] The regulations define impairments to include "[a]ny physiological disorder, . . . condition, cosmetic disfigurement, or anatomical loss affecting one" of the described biological "systems," as well as "[a]ny mental or psychological disorder, such as mental retardation, organic brain syndrome, emotional or mental illness, and specific learning disabilities."[100] The statute only defines major life activities by example, including "functions such as caring for oneself, performing manual tasks, walking, seeing, hearing, speaking, breathing, learning and working."[101] The Supreme Court has clarified that activities in this category must be "of central importance to daily life"[102] and "significan[t]."[103]

For a major life activity to be deemed "substantially limited," the plaintiff must come forward with evidence demonstrating that he or she is either unable to perform the activity or is "[s]ignificantly restricted as to the condition, manner or duration under which" the activity is performed.[104] In *Sutton v. United Air Lines, Inc.*,[105] the Supreme Court clarified that in evaluating whether this standard is satisfied, courts must consider all mitigating measures taken by plaintiffs to alleviate the effects of their impairments.[106] If the use of medication controls or corrects an individual's disorder, he or she generally will not be protected under the Act absent a showing of substantially limiting effects from the medication.[107] If the major life activity asserted is work, additional hurdles arise. Plaintiffs in this category must show that they are

---

Toyota Motor Mfg., Ky., Inc. v. Williams, 534 U.S. 184, 194 (2002); Sutton v. United Air Lines, Inc., 527 U.S. 471, 479 (1999). For an excellent discussion of the level of deference courts have given to the regulations in the wake of such questioning, *see* Eichhorn, *supra* note 14, at 209 (surveying cases and concluding that courts squarely presented with the issue of the regulations' validity are likely to conclude they are "valid legislative rules" or "interpretive rules that are entitled to a high degree of deference").

99. 42 U.S.C. § 12102(2)(A). Some plaintiffs in "regarded as" cases initially argued that they only had to demonstrate that employers regarded them as having an impairment rather than an impairment which substantially limits a major life activity. The Supreme Court has since clarified that all litigants, including those alleging a "regarded as" claim, must satisfy all three prongs of the disability definition in order to survive summary judgment. *See Sutton*, 527 U.S. at 489 ("[A] covered entity . . . must believe either that one has a substantially limiting impairment that one does not have or that one has a substantially limiting impairment when, in fact, the impairment is not so limiting.").

100. 29 C.F.R. § 1630.2(h)(1)-(2) (2004).

101. 29 C.F.R. § 1630.2(I) (2004).

102. *Toyota*, 534 U.S. at 197.

103. Bragdon v. Abbott, 524 U.S. 624, 638 (1998) ("'[T]he touchstone of determining an activity's inclusion under the statutory rubric [of a major life activity] is its significance.'" (quoting Bragdon v. Abbott, 107 F. 3d 934, 940 (1st Cir. 1997))).

104. 29 C.F.R. § 1630.2(j)(1)(i)-(ii) (2004).

105. 527 U.S. 471 (1999).

106. *Id.* at 482.

107. *Id.* at 482-83.

significantly restricted in performing "a class of jobs or a broad range of jobs" compared to similarly situated individuals, which is a significant challenge for most litigants.[108]

Establishing a protected disability, although necessary, is not sufficient in itself to gain coverage under the Act. The ADA only protects those individuals who also are considered "qualified" within the meaning of the statute.[109] To satisfy this requirement, plaintiffs must first show that they satisfy the prerequisites for the employment position, such as education or practical experience.[110] Next, plaintiffs must establish that they are able to perform the essential functions of the position "with or without reasonable accommodation."[111] In evaluating which functions are essential, courts examine "the fundamental job duties of the employment position,"[112] giving due consideration to the employer's judgment as to which functions fall within this definition.[113] If the plaintiff is unable to perform an essential job function, the court must also inquire whether a reasonable accommodation exists that would permit the employee to perform the function, such as job-restructuring or modification, the use of adaptive technology, or alterations to facilitate accessibility.[114] If such an accommodation is available, then the plaintiff is deemed a "qualified individual with a disability" within the meaning of the statute.[115]

B. Empirical Analysis

In the following section, this Article describes the results of empirical analysis undertaken to ascertain whether litigants with physical and mental impairments experience different success rates in establishing coverage under the ADA. Section II.B.1 describes the method of case selection and the process used to construct the sample. Section II.B.2 examines specific hypotheses that constitute the background of the study, as well as the methodology employed to test these hypotheses. Results are then presented in section II.B.3, followed by additional observations gleaned from the data in section II.B.4.

---

108. 29 C.F.R. § 1630.2(j)(3)(I) (2004) ("The inability to perform a single, particular job does not constitute a substantial limitation in the major life activity of working.").

109. 42 U.S.C. § 12111(8) (2000).

110. 29 C.F.R. § 1630.2(m) (2004).

111. 42 U.S.C. § 1211(8); 29 C.F.R. § 1630.2(m).

112. 29 C.F.R. § 1630.2(n) (2004).

113. 42 U.S.C. § 12111(8).

114. 42 U.S.C. § 12111(9) (2000).

115. 42 U.S.C. § 12111(8)-(9).

1. Data Summary

The study compiled a stratified sample of all Title I district and circuit court cases from the Third, Ninth, and Eleventh Circuit courts that were decided during the periods of 1994-1995 and 2002-2003 and that were available on Westlaw or LexisNexis. The three circuits were chosen because of their anecdotal reputations as "moderate," "liberal," and "conservative" jurisdictions respectively. Time periods before and after the *Sutton* decision were selected to evaluate whether this opinion influenced the success rates of plaintiffs in either or both groups. Following a review of all cases identified, the sample was limited to those cases in which the court explicitly discussed the "disabled" or "qualified" requirements, however briefly. The sample excluded cases in which the court assumed the existence of these elements or in which the parties did not dispute that they were satisfied. The study likewise excluded 12(b)(6) motions to dismiss and pre-trial motions that were not potentially dispositive. The resulting 131 claims were coded to indicate whether the alleged impairment was physical or mental,[116] whether the plaintiff successfully established a cognizable disability, and whether the plaintiff was deemed qualified for the employment position in question.[117] Information about the specific impairment alleged and the major life activity affected, as well as the type of disability claim, was likewise compiled.

2. Hypotheses & Methodology

As detailed earlier in this article, the study was motivated by the conventional wisdom that plaintiffs alleging mental impairments in Title I claims have more difficulty establishing coverage in the protected class than do plaintiffs alleging physical impairments. A chi-square test for equality of distributions was employed to test the validity of the primary statistical hypothesis that the rate at which mentally-impaired claimants establish class membership is equal to the rate at which physically-impaired claimants establish class membership.[118] The same test was used to evaluate whether

---

116. A small number of cases alleging mental disability were excluded (four cases involving only alcohol or drug addiction and one case involving a learning disability) to isolate psychiatric disorders that may provoke the stigmatization described *supra* Part I. Further, the unit of analysis employed was a particular claim as opposed to a case. Therefore, in a small number of cases where the court analyzed both mental and physical claims, these claims constitute independent records in the database. As a result, depending on what particular hypothesis is being tested, the sample size varies to some extent.

117. A list of the cases encompassing the 131 claims and their key attributes is on file with Tennessee Law Review.

118. The chi-square test provides the likelihood that a *difference* in proportions found in two or more groups could be found by chance when taking repeated samples. A small likelihood is an indication that there is some systematic difference between the groups. *See*

there are statistically significant differences in the ways these groups fail at class membership, i.e., due to failure to demonstrate a disability or due to a lack of qualification.

The sheer variability of the data concerning the specific impairments and major life activities alleged in each case, relative to the sample size, defies meaningful statistical analysis. However, some anecdotal observations related to these elements are offered below to shed additional light on the relationship between these requirements and plaintiffs' success rates in establishing class membership.

3. Results

Plaintiffs alleging mental impairments are only slightly less likely to establish coverage in the class protected by the ADA than those alleging physical impairments, and this difference is not statistically significant.[119] The sample reflected far more physical disability claims (103) than mental disability claims (28). Within the context of this sample, the mentally impaired plaintiffs prevailed in establishing class membership about 25% of the time, whereas the physically impaired plaintiffs prevailed about 29% of the time. For all practical purposes, these numbers reflect a statistical dead heat. See Table One.

			Established		Total
			No	Yes	
Type	Mental	Count	21	7	28
		% within Type	75.0%	25.0%	100.0%
		% within Established	22.3%	18.9%	21.4%
		% of Total	16.0%	5.3%	21.4%
	Physical	Count	73	30	103
		% within Type	70.9%	29.1%	100.0%
		% within Established	77.7%	81.1%	78.6%
		% of Total	55.7%	22.9%	78.6%
Total		Count	94	37	131
		% within Type	71.8%	28.2%	100.0%
		% within Established	100.0%	100.0%	100.0%
		% of Total	71.8%	28.2%	100.0%

Table One – Establishing Threshold Requirements, Mental vs. Physical Disabilities

GEORGE W. SNEDECOR & WILLIAM G. COCHRAN, STATISTICAL METHODS 76-79 (8th ed. 1989).

119. Table One, *Chi-Square* (1, N = 131) = .185, p = .667. This indicates that the difference in the two outcomes would be seen purely by chance approximately two-thirds of the time in repeated sampling.

There is a more meaningful difference, however, in the reasons courts identify for each group's failure to establish coverage in the protected class.[120] Whereas mentally-impaired plaintiffs lose most often because they are deemed unqualified by the court, physically impaired litigants lose most often because they do not establish that they are disabled within the meaning of the statute. See Table Two.

			Outcome			
			Not Disabled	Not Qualified	Survives Summary Judgment	Total
Type	Mental	Count	10	11	7	28
		% within Type	35.7%	39.3%	25.0%	100.0%
		% within Outcome	17.2%	30.6%	18.9%	21.4%
		% of Total	7.6%	8.4%	5.3%	21.4%
	Physical	Count	48	25	30	103
		% within Type	46.6%	24.3%	29.1%	100.0%
		% within Outcome	82.8%	69.4%	81.1%	78.6%
		% of Total	36.6%	19.1%	22.9%	78.6%
Total		Count	58	36	37	131
		% within Type	44.3%	27.5%	28.2%	100.0%
		% within Outcome	100.0%	100.0%	100.0%	100.0%
		% of Total	44.3%	27.5%	28.2%	100.0%

Table Two – Specific Outcomes, Mental vs. Physical Disabilities

In fact, focusing the analysis exclusively on those claims that fail to meet the threshold requirements exposes a difference between the two types of claims that achieves a moderate level of statistical significance.[121] Roughly two-thirds of mental impairment claims that fail to establish membership in the protected class do so because the plaintiff is found to be unqualified. Physical disability claims, on the other hand, fail to establish class membership roughly two-thirds of the time due to a failure to establish a protected disability. See Table Three.

120. These differences are still statistically insignificant. Table Two, *Chi-Square* (2, N = 131) = 2.528, $p = .283$. This indicates that the difference in the two outcomes would be seen purely by chance approximately 28% of the time in repeated sampling.

121. Table Three, *Chi-Square* (1, N = 79) = 3.668, $p = .055$. This indicates that the difference in the two outcomes would be seen purely by chance approximately 5% of the time in repeated sampling.

			Outcome		
			Not Disabled	Not Qualified	Total
Type	Mental	Count	5	8	13
		% within Type	38.5%	61.5%	100.0%
		% within Outcome	10.2%	26.7%	16.5%
		% of Total	6.3%	10.1%	16.5%
	Physical	Count	44	22	66
		% within Type	66.7%	33.3%	100.0%
		% within Outcome	89.8%	73.3%	83.5%
		% of Total	55.7%	27.8%	83.5%
Total		Count	49	30	79
		% within Type	62.0%	38.0%	100.0%
		% within Outcome	100.0%	100.0%	100.0%
		% of Total	62.0%	38.0%	100.0%

Table Three – Outcomes for Cases that Failed Threshold Requirements, Mental vs. Physical Disabilities

Restricting the sample to those cases where the disability threshold is clearly determinative exposes a highly significant difference in meeting the disability requirement prior to *Sutton* (1994-1995 in the sample) and after *Sutton* (2002-2003 in the sample).[122] While more than 60% of cases in the sample established a protected disability in 1994-1995,[123] only 28% were successful in 2002-2003. See Table Four. While the sample sizes in the current study are insufficient to analyze the effect of *Sutton* on mental cases and physical cases separately, purely descriptive statistics suggest that *Sutton* did not have a significant impact on the ability of plaintiffs alleging a mental impairment to establish class membership.[124] On the other hand, the impact on plaintiffs alleging physical impairments seems to be dramatic, with roughly 3 in 4 cases establishing class membership pre-*Sutton* and only 1 in 4 cases doing so post-*Sutton*. See Table Four.

---

122. Table Four, *Chi-Square* (1, N = 72) = 8.164, p = .004. This indicates that the difference in the two outcomes would be seen purely by chance less than 0.5% of the time in repeated sampling.

123. This closely corresponds to what has been reported elsewhere pre-*Sutton*. *See, e. g.*, Ruth Colker, *The Americans with Disabilities Act: A Windfall for Defendants*, 34 HARV. C.R.-C.L. L. REV. 99, 109 (1999) (reviewing ABA trial court data).

124. "Descriptive statistics are typically distinguished from inferential statistics. With descriptive statistics you are simply describing what is or what the data shows. With inferential statistics you are trying to reach conclusions that extend beyond the immediate data alone." William, M. K. Trochim, Descriptive Statistics, http://www.socialresearchmethods.net/kb/statdesc.htm (last visited Feb. 3, 2006).

			Established		
			Yes	No	Total
Date Group	1994-1995	Count	14	8	22
		% within Date Group	63.6%	36.4%	100.0%
		% within Established	50.0%	18.2%	30.6%
		% of Total	19.4%	11.1%	30.6%
	2002-2003	Count	14	36	50
		% within Date Group	28.0%	72.0%	100.0%
		% within Established	50.0%	81.8%	69.4%
		% of Total	19.4%	50.0%	69.4%
Total		Count	28	44	72
		% within Date Group	38.9%	61.1%	100.0%
		% within Established	100.0%	100.0%	100.0%
		% of Total	38.9%	61.1%	100.0%

Table Four – Establishing Disability Pre-*Sutton* (1994-1995) and Post-*Sutton* (2002-2003)

Overall, there is no statistical difference in these outcomes by circuit.[125] See Table Five.

			Established		
			No	Yes	Total
Circuit	11	Count	34	11	45
		% within Circuit	75.6%	24.4%	100.0%
		% within Established	39.5%	33.3%	37.8%
		% of Total	28.6%	9.2%	37.8%
	3	Count	42	17	59
		% within Circuit	71.2%	28.8%	100.0%
		% within Established	48.8%	51.5%	49.6%
		% of Total	35.3%	14.3%	49.6%
	9	Count	10	5	15
		% within Circuit	66.7%	33.3%	100.0%
		% within Established	11.6%	15.2%	12.6%
		% of Total	8.4%	4.2%	12.6%
Total		Count	86	33	119
		% within Circuit	72.3%	27.7%	100.0%
		% within Established	100.0%	100.0%	100.0%
		% of Total	72.3%	27.7%	100.0%

Table Five – Establishing Threshold Requirements, by Circuit

125. Table Five, *Chi-Square* (2, N = 119) = .512, p = .774. This indicates that the difference in the two outcomes would be seen purely by chance approximately three-fourths of the time in repeated sampling.

4. Other Observations

The cases in the sample demonstrated a great deal of diversity in the types and combinations of impairments identified, the types and combinations of major life activities asserted, and the disability prong employed to allege these elements. While this complexity makes meaningful statistical analysis practically impossible, some aggregation and classification of cases, admittedly not always exact, nevertheless reveals some interesting descriptive statistics.

The cases in the aggregate sample were evenly split between plaintiffs alleging an actual disability (33%) and those alleging that they were regarded as disabled by their employers (33%). Those plaintiffs in the former category were more successful at establishing a protected disability (60%) than those in the latter category (40%). Hybrid strategies alleging both types of disability were alleged in 33% of the cases but were successful only 25% of the time.

Plaintiffs in a surprising number of mental impairment cases alleged that they suffered from depression, either alone or in combination with other disorders. About 25% of the psychiatric cases in the sample identified depression as the exclusive impairment, while another 25% of these cases alleged depression in combination with other disorders, such as post-traumatic stress or anxiety. The success rate of plaintiffs in establishing a disability on the basis of depression (50%) exceeded that of plaintiffs alleging other mental impairments (29%). Those plaintiffs who alleged depression in combination with other impairments, however, were the most successful in establishing a cognizable disability (75%).

Despite the Supreme Court's skepticism over the viability of work as a major life activity,[126] a significant number of plaintiffs in the sample (40%) alleged it as the exclusive major life activity for consideration, while another 33% alleged work in combination with some other activity. The remaining 27% of the sample identified a wide variety of non-work-related activities. Despite the prevalence of such claims, they were rarely successful. Cases alleging work as the exclusive activity successfully established a disability only 17% of the time, while cases alleging work in combination with some other activity were slightly more successful, with 40% establishing a protected disability. These results are consistent with scholarly works indicating that alleging work alone is rarely a winning strategy for any plaintiff.[127]

---

126. *See* Toyota Mfg., Ky., Inc. v. Williams, 534 U.S. 184, 200 (2002) ("Because of the conceptual difficulties inherent in the argument that working could be a major life activity, we have been hesitant to hold as much, and we need not decide this difficult question today."); Sutton v. United Airlines, Inc., 527 U.S. 472, 492 (1991) (questioning whether work is an appropriate major life activity and "[a]ssuming without deciding that working is a major life activity").

127. *See, e.g.*, Steven S. Locke, *The Incredible Shrinking Protected Class: Redefining the*

IV. ANALYSIS

The empirical study discussed above did not reveal the anticipated results. Contrary to the weight of scholarly opinion, individuals alleging psychiatric impairments did not experience heightened difficulty in establishing either a protected disability or membership in the protected class. The question thus becomes what may account for the discrepancy between the widely held belief and these findings. It is plausible that the results would be different if the sample were expanded, and this would prove a fruitful area for future research and inquiry. Alternatively, the results may reflect society's growing acceptance of individuals with mental illness as a result of public interest campaigns and expanding knowledge about the causes of psychiatric disorders. This conclusion seems unlikely, however, given the numerous indicators reflecting that stigma remains widespread discussed earlier in this Article.[128] It may be that the results actually reflect the continuing presence of mental illness stigma, although in more subtle and indirect ways than anticipated. The remainder of this section discusses two potential explanations for the unexpected results and proposes areas for future research.

A. The Problem of Disclosure: A Cost-Benefit Calculus

The success of those with mental impairments in establishing protected disabilities and coverage in the class protected by the ADA may perversely reflect the high degree of risk that attaches to self-identification of this type of disability in the workplace. The stereotypes associated with mental illness are significantly more negative and more pervasive than those associated with physical disabilities.[129] As a result, there are very real disincentives to disclosure of a psychiatric disorder at the time of hire or during the employment relationship. It is plausible that only those with the highest need for accommodation, who correspondingly experience the most severe disorders, will risk identification to seek assistance. These same individuals are most likely to establish an impairment which substantially limits a major life activity in litigation.

The costs of disclosure when applying for a job are readily apparent. Numerous studies have identified employers' reluctance to associate with those they consider mentally ill.[130] One study concluded that employers prefer to hire an individual who has spent time in prison over an individual with

---

Scope of Disability Under the Americans with Disabilities Act, 68 U. COLO. L. REV. 107, 135 (1997) ("[C]laims in which plaintiffs have only alleged a substantial limitation in the major life activity of working have been almost universally rejected.").

128. *See supra* Part I.

129. *See supra* Part I.

130. *See* Campbell, *supra* note 4, at 137 (describing various studies).

mental instability.[131] Yet another revealed that 60% of personnel directors "would never choose an individual with mental illness for an executive job," compared with just 3% who would not hire an individual with diabetes.[132] Employers are uneasy about the behavior individuals with psychiatric disorders may display on the job if hired and the subsequent disruption in the workplace.[133] Such findings are mirrored by the 31% of applicants in a recent study who reported that they had been turned down for a job they were qualified for after revealing a mental illness.[134] Some vocational rehabilitation counselors, to avoid such concerns, actually encourage their clients to hide prior hospitalizations or to devise strategies for covering gaps in employment caused by mental illness.[135]

The ADA's prohibition on employers asking medical questions during the pre-hire stage recognizes this bias and provides some protection during the initial interview phase.[136] Because psychiatric disorders are often difficult to detect by third parties, interviewees with mental impairments, as opposed to many of their physically impaired counterparts, can more readily keep their disabilities hidden until a job offer is extended. Once hired, however, employees are unable to assert their rights under the ADA and secure assistance in the workplace without disclosing the nature of the underlying impairment.[137] The costs of disclosure, even after establishing a successful work history for an employer, are again potentially steep. Despite the law's

---

131. *Id.* (citing James N. Colbert et al., *Two Psychological Portals of Entry for Disadvantaged Groups*, 34 REHABILITATION LITERATURE 194, 199 (1973)).

132. Otto F. Wahl, TELLING IS RISKY BUSINESS: MENTAL HEALTH CONSUMERS CONFRONT STIGMA 84-85 (1999).

133. *See* Edward Diksa & E. Sally Rogers, *Employer Concerns About Hiring Persons with Psychiatric Disability: Results of the Employer Attitude Questionnaire*, 40 REHABILITATION COUNSELING BULL. 31, 42 (1996).

134. Wahl, *supra* note 132, at 80.

135. *See* Beth Angell et al., First-Person Accounts of Stigma, *in* ON THE STIGMA OF MENTAL ILLNESS 73 (Patrick W. Corrigan, ed., 2005) (noting studies suggest that individuals with histories of mental illness "may creatively construct a work history that conceals gaps in their employment record."); Campbell, *supra* note 4, at 140.

136. *See, e.g.*, 42 U.S.C. § 12112(d)(2) (2004); 29 C.F.R. § 1630.13 (2004).

137. *See* Angell, *supra* note 135, at 74 (discussing the "double-edged sword" faced by mentally ill employees who fail to disclose their impairments but need accommodations in the workplace). Employers need only make "reasonable accommodations to the known . . . . limitations of an otherwise qualified individual with a disability." 42 U.S.C. § 12112(b)(5)(A) (2000). Employees experiencing disparate treatment, moreover, must ultimately show in litigation that the employer discriminated "because of" a disability. 42 U.S.C. § 12112(a) (2000). This standard is hard to meet in the absence of clear evidence that the employer knew or should have known of the disability. *Cf.* Ann C. Hodges, *The Americans with Disabilities Act in the Unionized Workplace*, 48 U. MIAMI L. REV. 567, 611 (1994) ("[E]stablishing discrimination [by a union] will be extraordinarily difficult if the individual cannot prove the union's knowledge of the disability."). Self-disclosure most readily establishes such knowledge. *Id.*

anti-discrimination mandate, individuals with mental illness continue to report that they are demoted or even fired once a psychiatric diagnosis becomes common knowledge in the workplace.[138] Those who retain their positions are likely to be adversely impacted in other ways. Individuals labeled as mentally ill on the job are more likely to earn less income and be underemployed than similarly situated impaired individuals who have not yet been labeled.[139] In one study, employers who were asked to rate employees for promotion purposes viewed employees with depression as "less worthy of a salary increase, less productive, and less dependable," less confident, less able to be accepted by co-workers, and more negatively responsive to criticism than equally qualified non-disabled peers.[140] The study concluded that the perceived cause of the employee's disability was directly linked to the diminished recommendation for promotion.[141] Because of the persistent link between character defects and mental illness discussed earlier, it is likely that the study's findings extend beyond depression to encompass all psychiatric disorders.[142]

Disclosure likewise may damage relationships with co-workers and supervisors. The revelation of a diagnosis may lead to harassment, teasing, and isolation even in a formerly comfortable workplace.[143] Fears of dangerousness, unpredictability, and poor behavior by the labeled employee are common. Employees who secure a requested accommodation may face even more hostility, as coworkers respond with "ambient jealousy" to what they regard as special treatment for malingerers.[144] Such reactions are especially problematic in light of the heightened difficulties employees with psychiatric disabilities often have establishing interpersonal relationships with coworkers in the first instance.[145]

Faced with the prospect of either not securing employment or limiting advancement opportunities upon disclosure, only those individuals most in need of the ADA's protection are likely to judge disclosure worth the risk. Employees and applicants must weigh the advantage of securing protection against both the animus toward mental illness and the increasing uncertainty

---

138. Wahl, *supra* note 132, at 84-85.

139. *See* Link, *supra* note 43, at 202, 209-10.

140. James E. Bordieri et al., *Work Life for Employees With Disabilities: Recommendations for Promotion*, 40 REHABILITATION COUNSELING BULL. 181, 186 (1997).

141. *Id.* at 188-89 ("Specifically, candidates seen as personally responsible for their disabilities were given lower promotion recommendations than were candidates whose disabilities were attributed to external factors.").

142. *See supra* notes 45-56 and accompanying text.

143. For a revealing discussion of first-hand stories in this regard, see Wahl, *supra* note 132, at 85-87.

144. Owen, *supra* note 9, at 94, 118.

145. David E. Drehmer & James E. Bordieri, *Hiring Decisions for Disabled Workers: The Hidden Bias*, 30 REHABILITATION PSYCHOL. 157, 158 (1985) (discussing the results of studies by several authors).

of whether disclosure will bring legal protection or assistance. Employees may understand that the promise of protection will be illusory if a judge deems their impairment insufficient for protection under the ADA.[146] Those who weigh the calculus in favor of disclosure, therefore, may be those with the more severe impairments who have the greatest likelihood of making the requisite legal showing of a protected disability in litigation. If correct, this theory directly counters the popular image of the ADA plaintiff, who is often portrayed as an unimpaired malingerer who asserts frivolous problems in order to lighten workloads and secure favorable treatment in the workplace.[147]

This theory, although helpful in explaining the results of the study, has its limitations. It best describes those cases involving an employer's failure to accommodate an actual disability, where the severity of the need to disclose may correlate most closely to the severity of the impairment, and may be less helpful in explaining class coverage in disparate treatment and "regarded as" cases.[148] It is worth noting, moreover, that psychiatric disability claims are the second largest category of EEOC complaints, which may indicate that disclosure occurs more frequently than the low levels of litigated claims suggest.[149] Further research, therefore, is necessary to test the accuracy and strength of this theory.

---

146. Owen, *supra* note 9, at 121-22. Discussing the results of her small study, Owen concludes that "the ADA seems to be regarded by the workers in this study as intriguing and sometimes relevant to their circumstances, but risky to embrace in today's political economy." *Id.* at 169.

147. *See, e.g.*, Cary LaCheen, *Achy Breaky Pelvis, Lumber Lung and Juggler's Despair: The Portrayal of the Americans with Disabilities Act on Television and Radio*, 21 BERKELEY J. EMP. & LAB. L. 223, 227 (2000)

> [T]hose with hidden disabilities who did not reveal their conditions in the past but choose to do so when they invoke the ADA are regarded with strong suspicion, and the media coverage often suggests or implies that the only plausible explanation for this is fraud, [sic] and a desire to avoid work.

Id. Cf. Parikh, *supra* note 16, at 743 ("The lack of outward warning signs make the [mental] illness inherently more suspect, as many believe it could easily, and conveniently, be faked.").

148. In disparate treatment cases, the need to disclose may be more dependent on the egregiousness of the employer's behavior than the severity of the plaintiff's impairment. In "regarded as" cases, there may be no impairment, severe or otherwise, to disclose. *See supra* note 99.

149. Owen, *supra* note 9, at 173. At least one national survey suggests that 86.6% of the professionals and managers with psychiatric conditions disclose their conditions while employed. Marsha Langer Ellison et al., *Patterns and Correlates of Workplace Disclosure Among Professionals and Managers with Psychiatric Conditions*, 18 J. VOCATIONAL REHABILITATION 3, 7 (2003). One small qualitative study, however, recently found lower rates of disclosure among people with psychiatric disorders generally. *See* Susan G. Goldberg et al., *The Disclosure Conundrum: How People With Psychiatric Disabilities Navigate Employment*, 11 PSYCHOL. PUB. POL'Y & L. 463, 488 (2005) ("The high rate of disclosure [in Ellison et. al's study] may be due to the higher proportion of participants in [that] study who worked in the field of mental health (40%) and for whom disclosure may have been a necessity.").

B. The Problem of Selection

An alternate explanation for the study's results may lie in the source from which the sample was constructed. As detailed above, the sample consists of published and unpublished decisions that were decided during the selected time periods and were available in Westlaw and LexisNexis. Significant evidence suggests, however, that litigated claims do not reflect a representative sample of the entire body of disputes; the vast majority of disputes are settled before ever going to trial.[150] It may be that the study's reliance on litigated claims inadvertently masks the increased difficulty plaintiffs with psychiatric impairments actually experience in establishing coverage under the ADA.

Building upon economic theories of suit and settlement,[151] Priest and Klein have proposed that claims with quality that deviate much from the decision standard, above or below, will naturally settle.[152] Generally counsel will accurately assess the value of very weak and very strong claims, which are dropped or directed towards non-judicial resolution. The remaining litigated claims would be only the close calls; so close, in fact, that judgments would be about evenly split between the two sides, and the likelihood of success would tend towards 50%.[153]

---

150. *See generally* George L. Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J. LEGAL STUD. 1-6 (1984).

> Virtually all systematic knowledge of the legal system derives from studies of appellate cases. Appellate cases, of course, provide the most direct view of doctrinal developments in the law. . . . But this doctrinal information discloses very little about how legal rules affect the behavior of those subject to them or affect the generation of legal disputes themselves.

Id. at 1. Trial occurs only if $P_p - P_d > (C - S)/J$ where P_p is the plaintiff's estimated probability of success, P_d is the defendant's estimated probability of success, C is the parties' collective trial costs, S is the parties' collective settlement costs, and J is the prospective judgment upon success. Joel Waldfogel, *Reconciling Asymmetric Information and Divergent Expectations Theories of Litigation*, 41 J.L. & ECON. 451, 453-55 (1998).

151. *See generally*, John P. Gould, *The Economics of Legal Conflicts*, 2 J. LEGAL STUD. 279 (1973); William M. Landes, *An Economic Analysis of the Courts*, 14 J.L. & ECON. 61 (1971); Richard A. Posner, *An Economic Approach to Legal Procedure and Judicial Administration*, 2 J. LEGAL STUD. 399 (1973).

152. *See* Priest & Klein, *supra* note 150, at 4-5.

153. *Id.* ("[W]here the gains or losses from litigation are equal to the parties, the individual maximizing decisions of the parties will create a strong bias toward a rate of success for plaintiffs at trial or appellants at appeal of 50 percent regardless of the substantive standard of law."); Waldfogel, *supra* note 150, at 454.

> In this model, settlement acts as a two-sided filter on the population of filed cases. If a case has true quality far above or below the decision standard, it is unlikely that the parties will disagree sharply about the plaintiff's prospects at trial. The cases most likely to go to trial are those with true quality near the decision standard, giving rise to the tendency toward 50% plaintiff victories at trial. As the degree of error in parties' case quality

Clearly, the predicted 50% split is not ordinarily observed in empirical analysis of litigation, nor is it the result found in this Article. Scholars have offered a number of explanations for this divergence, including differences in stakes between the parties[154] and asymmetric information,[155] both of which could predict plaintiff success rates less than 50%. The key point, however, is that holding costs and prospective judgments constant,[156] cases would be expected to converge to the same likelihood of success even if the decision standards were not the same.[157] Thus, even if judges apply a more difficult

---

estimates declines, or if the relative gains from trade . . . rise, the trial filter becomes finer. With more accurate estimates of case quality, higher trial costs, or smaller judgments, only cases closer to the decision standard go to trial, and the ensuing plaintiff win rates at trial converge to 50%.

Id.

154. Priest & Klein, *supra* note 150, at 26. When there are differences in stakes, as in the case when the defendant has greater precedential stake than does the plaintiff, differences in likelihoods of success at trial may result. "As a consequence, relatively more disputes with likely plaintiff [outcomes] will be settled and relatively more disputes with likely defendant [outcomes] will be litigated. Observing only litigated cases, defendant verdicts will be greater than 50%." *Id.*

155. Information asymmetry is based on the idea that one party has more information (e.g., knows more about what actually happened, has more experience with this type of litigation, or has more experienced counsel) about the likely outcome of the trial. *See generally* Robert H. Gertner, *Asymmetric Information, Uncertainty, and Selection Bias in Litigation*, 1993 U. CHI. L. SCH. ROUNDTABLE 75; Keith N. Hylton, *Asymmetric Information and the Selection of Disputes for Litigation*, 22 J. LEGAL STUD. 187 (1993); Waldfogel, *supra* note 150 at 452.

In the [asymmetric information] theory one party knows the probability that the plaintiff will win at trial, while the other party knows only the distribution of plaintiff victory probabilities. When the defendant is better informed, the (uninformed) plaintiff makes a settlement offer, and it is accepted by informed defendants who face a relatively high expected liability at trial. The defendants proceeding to trial, on the other hand, are those who correctly expect to win. The selection of cases for trial is thus one-sided, and the plaintiff win rate at trial is systematically below the fraction of plaintiff winners in the filed pool.

Id.

156. The decision to bring a case depends on the relationship of the predicted likelihood of success to costs and prospective judgments. For the plaintiff,

$$E(C) = P_p(J_p - C_p) + (1 - P_p)(-C_p)$$

where $E(C)$ is the expected value of litigating the case, P_p is the predicted likelihood of success, J_p is the prospective judgment, and C_p is litigation costs. Setting $E(C)$ to zero assuming that the plaintiff will litigate any case with a positive expected value,

$$P = C_p / J_p .$$

157. Priest & Klein, *supra* note 150, at 17 ("[D]isputes lying close to the decision standard are more likely to be litigated than disputes lying far from the decision standard. . . . The

threshold to claims alleging mental impairments, such heightened scrutiny may have no effect at all on the claims' likelihood of success.

An analogy familiar to law professors may help to illuminate this hypothesis. Suppose that, facing one hundred exams to grade, the goal is simply to sort them into two groups: those that earn a B or above and those that earn a C or below. Most papers will clearly belong in one group without much consideration. These papers are analogous to the cases that settle. Only a handful of papers close to the dividing line between the two grades will need more exacting scrutiny. These papers, analogous to litigated cases, are likely to be very close to the professor's dividing line between grades, or in law and economics parlance, the decision standard. It is plausible to assume that of this close group, roughly half the papers will earn B's and half C's.

Assume now, however, that the professor wishes to divide the papers between those that earn A's and those that earn B's and below, thus applying a heightened decision standard. The process is largely the same in that there are clear A's, clear "not-A's," and a small group of papers that need more attention to decide. These papers similarly cluster around the new standard. If close enough, approximately one-half of these will successfully earn A's and the other half will end up in the less desirable stack. Moving the decision standard higher for this group of papers should have no effect on the likelihood that a paper from the close scrutiny group will end up in the top stack as opposed to the lower stack.

Applying these theories to the current research findings, it would be reasonable to expect that in cases when the issue of disability or qualification is clear, these issues may not be disputed (a result occasionally seen when constructing the sample) and this fact may impact the appraisal of the overall value of the case. Given that the defendant either had a larger stake in the outcome, which might be the case when precedent is established in a particular workplace, or enjoyed an informational advantage, which would be a reasonable expectation given that many defendants are ADA repeat players,[158] the success rate for plaintiffs would converge on something less than half, such as the roughly 25% found here.

Most importantly, assuming that the cost of litigating a mental disability case is approximately equal to the cost of litigating a physical disability case, and assuming that the prospective judgments in the two cases are roughly the

---

proportion of plaintiff victories in litigated disputes will approach 50 percent regardless of the position of the decision standard with respect to the underlying distribution of disputes.").

158. Indeed, the low success rate among ADA plaintiffs maybe attributed to a defense bar that significantly out-lawyers the plaintiff bar. *See generally* Jefferey A. Van Detta & Dan R. Gallipeau, *Judges and Juries: Why are so Many ADA Plaintiffs Losing Summary Judgment Motions, and Would They Fare Better Before a Jury? A Response to Professor Colker*, 19 REV. LITIG. 505 (2000) (arguing that many ADA plaintiffs do not survive summary judgment due to poor lawyering); Wendy Wilkinson, *Judicially Crafted Barriers to Bringing Suit Under the Americans with Disabilities Act*, 38 S. TEX. L. REV. 907 (1997) (arguing that many ADA cases that do not survive summary judgment have been poorly pleaded).

same, which is reasonable given that damages in both case types would be based primarily on similar employment-related factors, the fact that a mental disability case may have a decision standard that is more arduous than that of a physical disability case would not affect the likelihood of success.[159] The difference in decision standards would affect the number of each type of case litigated. Given a more difficult decision standard for establishing disability and qualification in mental disability cases, plaintiffs and their counsel would bring higher quality cases as compared to physical disability case in order to meet this elevated standard. Returning to the exam analogy, while the success rates in the close scrutiny group related to the A's/not-A's standard may be quite similar to the close scrutiny group related to the B's-and-above/C's-and-below standard, one ordinarily would find that there are significantly fewer A's/not-A's close calls as compared to B's-and-above/C's-and-below close calls precisely because the standard is higher. Similarly, fewer mental disability cases would be brought than physical disability cases if the standard were higher. That is exactly what is found in this study.[160]

This theory also suggests some opportunities for further research. Three factors require further investigation to be confident that selection bias explains the results of the study. Validating the equivalence of the selection models requires showing that the costs of litigating a mental disability case are approximately the same as those for a physical disability case, and that the prospective judgments in the two types of cases are roughly equal. In addition, it would be helpful to find some evidence that the nature and prevalence of the discrimination that is the source of the disputes and the distribution of potential cases are comparable for the two types of cases.

V. BRIDGING THE GAP: ALLEVIATING THE STIGMA OF MENTAL ILLNESS

The continuing impact of mental illness stigma on a plaintiff's ability to establish threshold coverage under the ADA is unclear in this study. The comparatively similar success rates between plaintiffs with physical impairments and those with mental impairments may suggest that stigma plays less of a role in judicial decisionmaking than hypothesized by scholars. Alternatively, it may reflect the high degree of risk attached to the disclosure of psychiatric disorders in the workplace, bias in the sample tested, or some other subtle form of stigma. As identified throughout this Article, further research is necessary to test which conclusion is likely to best explain what is

159. What influences the particular outcomes of litigated cases is not the decision standard, but the difference in the plaintiff's and defendant's estimation of the quality of a given case vis-à-vis the decision standard. *See supra* notes 150-57 and accompanying text. Those cases close to the higher mental disability decision standard would be influenced in the same way by divergent expectations and asymmetrical information as those case close to the lower physical disability decision standard; they would converge on the same success rate.

160. Of 131 claims considered, only 28 claimed a mental disability.

actually happening in the litigation process.

More immediate insight is gained from the disparate rationales offered by courts in denying class membership to physically and mentally impaired plaintiffs. The ability of plaintiffs with psychiatric disorders to establish a protected disability at least as readily as, if not more readily than, plaintiffs with physical impairments suggests that judges do not routinely endorse the stereotype that mental illness is easily faked, and that most who pursue such claims are malingerers. The relatively high number of plaintiffs with psychiatric disorders in the sample who failed to establish that they are qualified, however, seems to reflect the continuing belief that mental illness automatically renders individuals incompetent and incapable of performing in the workplace. Whether these failure rates also reflect continuing concerns with the dangerousness of group members or beliefs about their morally-defective character remains uncertain.

In either case, such results suggest that mental health advocates may be well advised to direct public education efforts toward changing attitudes about the "normalcy" and competence of group members rather than toward establishing mental illness as a legitimate health problem. The implications for litigants may be more complicated. One could argue that advocates for individuals with psychiatric disorders would do well to focus proportionately more attention on the high threshold applied to the qualified requirement than on the difficulty of establishing a cognizable disability under the statute. The inter-relatedness between these two elements, however, belies this approach. As scholars have noted, plaintiffs seeking the ADA's protection are given the herculean task of establishing a substantially-limiting impairment while at the same time demonstrating sufficient qualifications for the employment position.[161] This catch-22 has become more profound as a result of Supreme Court decisions' raising the bar on the definition of disability.[162] It may be that the only effective way to resolve this tension is to change the statutory definition of disability altogether.[163] Given the lack of Congressional reaction to the Supreme Court decisions, however, such action seems unlikely at best for the foreseeable future.

In the final analysis, more empirical research is needed to understand the

---

161. *See, e.g.*, Hensel, *supra* note 16, at 1188.

162. *See, e.g.*, Sutton, 527 U.S. at 482.

163. *See, e.g.*, Chai Feldblum, *Definition of Disability Under Federal Anti-Discrimination Law: What Happened? Why? And What Can We Do About It?*, 21 BERKELEY J. EMP. & LAB. L. 91, 162 (2000) (proposing legislative changes to the ADA, including "amend[ing] the definition of disability to mean a physical or mental impairment.").

ways in which mental illness stigma may influence litigation outcomes under the ADA. Armed with such information, plaintiffs with psychiatric disorders will more readily secure the full and equal protection that the ADA was intended to offer to all people with disabilities.

H-2A GUEST WORKER PROGRAM: EMPLOYER CERTIFICATION PROCESS IN NEED OF A CHANGE

I. INTRODUCTION

The Immigration and Nationality Act of 1952[1] allows alien workers to work temporarily in the United States for agricultural employers.[2] Foreign workers, or "H-2A workers," who enter the United States under the "H-2A" guest worker program,[3] are issued short-term visas to perform work of a "temporary or seasonal nature."[4] In recent years, the number of H-2A applications submitted to the Department of Labor has risen, and the demand for H-2A workers continues to grow as the agricultural industry increasingly relies on this class of workers as their primary labor source.[5]

To hire foreign workers, U.S. agricultural employers must seek approval from the Employment and Training Administration of the United States Department of Labor.[6] Approval will be granted if the agricultural employer demonstrates that (1) no "sufficient" U.S. workers are "able, willing, and qualified, [or] . . . available at the time and place needed" to provide the desired services and (2) employment of H-2A guest workers "will not adversely affect the wages and working conditions of workers in the United States similarly employed."[7]

At first glance, these restrictions on agricultural employers in the current

---

1. Act of June 27, 1952, ch. 477, 66 Stat. 163 (codified as amended at 8 U.S.C. §§ 1101-1537 (2000)).

2. Immigration and Nationality Act, 8 U.S.C. §§ 1101(a)(15)(H)(ii)(a), 1188(a)(1) (2000). The Act also permits foreign workers to come to the United States "to perform [non-agricultural] temporary service or labor" if employers cannot find unemployed U.S. workers to fill these positions. *Id.* § 1101(a)(15)(H)(ii)(b). These workers, named after the relevant section of the statute, are known as "H-2B workers." *See* U.S. Dep't of Labor, Employment & Training Admin., H-2B Certification for Temporary Nonagricultural Work, http://workforcesecurity.doleta.gov/foreign/h-2b.asp (last visited Jan. 8, 2006).

3. *See* 8 U.S.C. § 1188(a)(1), (i)(2).

4. *Id.* § 1101(a)(15)(H)(ii)(a); *id.* § 1201(a)(2), *amended by* 8 U.S.C. § 1201(a)(1)(B) (Supp. II 2002). "Temporary or seasonal nature" as used in the statute "means employment performed at certain seasons of the year, usually in relation to the production and/or harvesting of a crop, or for a limited time period of less than one year when an employer can show that the need for the foreign worker(s) is truly temporary." U.S. Dep't of Labor, Employment & Training Admin., H-2A Certification, http://workforcesecurity.doleta.gov/foreign/h-2a.asp (last visited Jan. 8, 2006).

5. *See* Josh McDaniel, *The H-2A Program: Hiring Temporary Foreign Workers for Farm Labor* (2003), http://www.aces.edu/pubs/docs/A/ANR-1244. The majority of H-2A workers are located in the southeastern states. *Id.*

6. *See* 20 C.F.R. § 655.101 (2005).

7. *See* 8 U.S.C. § 1188(a)(1).

H-2A guest worker program appear to benefit and protect U.S. workers. In reality, however, the program is a flawed system that does not adequately protect our "able, willing, qualified, and . . . available" U.S. workforce. Agricultural employers often abuse the program and exploit vulnerable foreign workers; this abuse is enabled by the Department of Labor's failure to ensure that agricultural employers meet all of the requirements and conditions set forth in the federal regulations.

Nevertheless, the H-2A guest worker program is vital to the agricultural industry. This comment discusses the current state of and issues surrounding the H-2A employer certification process. In addition, it addresses the need for Congress to protect the integrity of the H-2A system by enacting new legislation to remedy the abuses that plague the current program. Finally, this comment provides a potential solution to employer abuse of the H-2A guest worker program: the creation of a formal avenue of appeal that would allow U.S. workers to challenge the Employment and Training Administration's decision to consider an agricultural employer's application to hire foreign guest workers.

II. DEFINITIONS

To fully comprehend the problems with and the abuses of the H-2A guest worker program, one must first understand the program itself. Therefore, an examination of the definitions assigned to the key terms used in the H-2A guest worker program is critical.

A. "Agriculture"

For purposes of the H-2A program, "agriculture" is defined either by using the Internal Revenue Code's definition of "agricultural labor" or the Fair Labor Standards Act of 1938's simple definition of "agriculture."[8] The Internal Revenue Code defines "agricultural labor" as "all service performed on a farm . . . in connection with cultivating the soil, . . . raising or harvesting any agricultural . . . commodity, including the . . . management of livestock, bees, poultry, and fur-bearing animals and wildlife."[9] This definition also encompasses services performed on a farm "in connection with the operation . . . [and] maintenance of such farm," the harvesting of agricultural commodities as defined by the Agricultural Marketing Act,[10] all services "in connection with the ginning of cotton," and services related to the growth, storage, packing, and delivering of agricultural commodities.[11] The term "farm" within the definition includes "stock, dairy, poultry, fruit, fur-bearing

8. 20 C.F.R. § 655.100(c)(1).
9. I.R.C. § 3121(g)(1) (2000).
10. Agicultural Marketing Act, 12 U.S.C. § 1141j(g) (2000).
11. I.R.C. § 3121(g)(2)-(4).

animal, and truck farms, plantations, ranches, nurseries, ranges, greenhouses or other similar structures used primarily for the raising of agricultural . . . commodities."[12]

The Fair Labor Standards Act of 1938 defines "agriculture" as

> farming in all its branches and among other things includes the cultivation and tillage of the soil, dairying, the production, cultivation, growing, and harvesting of any agricultural . . . commodities . . . , the raising of livestock, bees, fur-bearing animals, or poultry, and any practices (including any forestry or lumbering operations) performed by a farmer or on a farm as an incident to or in conjunction with such farming operations, including preparation for market, delivery to storage or to market or to carriers for transportation to market.[13]

B. "Temporary or Seasonal Nature"

H-2A guest workers can only gain entry into the United States to perform agricultural services that are "of a temporary or seasonal nature."[14] According to federal regulations, "temporary or seasonal" labor constitutes work that is "usually of short duration" or "is of the kind exclusively performed at certain seasons or periods of the year and which, from its nature, may not be continuous or carried on throughout the year."[15] This definition does not encompass "supervisory employee[s] . . . employed by . . . specific agricultural employer[s] . . . on a year round basis."[16]

C. "Temporary Worker"

H-2A guest workers enter the United States on a *temporary* basis without the intention of becoming permanent residents.[17] Federal regulations define "temporary" as a time period of "less than one year."[18] Therefore, H-2A workers are not permitted to work and live in the United States for more than one year and do not acquire any right to remain in the United States once their visas expire. In fact, H-2A workers never have the opportunity to become immigrants under the current H-2A guest worker program even though many

---

12. *Id.* § 3121(g).
13. Fair Labor Standards Act of 1938, 29 U.S.C. § 203(f) (2000).
14. 8 U.S.C. § 1101(a)(15)(H)(ii)(a) (2000).
15. 29 C.F.R. § 500.20(s)(1)-(2) (2005). For example, an apple farm's picking season may last for three months. Any services performed on the farm would qualify as "temporary or seasonal" because they are only needed for a period of three months and could not be performed year round.
16. *Id.* § 500.20(s)(3).
17. 8 U.S.C. § 1101(a)(15)(H)(ii)(a).
18. 20 C.F.R. § 655.100(c)(2)(iii).

of them return to the United States year after year with temporary visas.[19]

III. TEST FOR THE ADMISSION OF H-2A WORKERS

Many cases have established that the purpose of the complex statutory and regulatory provisions governing the H-2A guest worker program is to protect U.S. agricultural workers.[20] U.S. workers—individuals who are legally permitted to work in the United States[21]—are given a preference over foreign guest workers for agricultural jobs in this country.[22] Federal regulations require that two conditions be met before the Department of Labor may authorize an agricultural employer to hire H-2A guest workers.

First, agricultural employers must demonstrate that there is an insufficient number of U.S. workers "who are able, willing, and qualified, and who will be available at the time and place needed, to perform the labor or services involved in the petition."[23] Employers may impose certain job qualifications for U.S. workers only if the qualifications are essential for performing the requested services.[24] Any qualifications must both be approved by the Regional Administrator and be among "the normal and accepted qualifications required by non-H-2A employers in the same or comparable occupations."[25] This essentially precludes agricultural employers from imposing trivial job qualifications on U.S. workers solely to prevent them from being qualified for positions.

Second, agricultural employers must establish that the employment of H-2A workers "will not adversely affect the wages and working conditions of

---

19. The current legislation provides no avenue for H-2A workers to gain permanent resident status in the United States. However, a widely supported bill presently under consideration by Congress, known as the Agricultural Job Opportunities, Benefits, and Security Act of 2005, would permit H-2A guest workers to attain permanent residency. *See* Agricultural Job Opportunities, Benefits, and Security Act of 2005, S. 359, 109th Cong. (2005).

20. *See, e.g.*, Alfred L. Snapp & Son, Inc. v. Puerto Rico *ex rel.* Barez, 458 U.S. 592, 596 (1982); Hernandez Flecha v. Quiros, 567 F.2d 1154, 1155 (1st Cir. 1977); Elton Orchards, Inc. v. Brennan, 508 F.2d 493, 500 (1st Cir. 1974).

21. 20 C.F.R. § 655.100(b).

22. *See* Alfred L. Snapp & Son, 458 U.S. at 596.

23. 8 U.S.C. § 1188(a)(1)(A) (2000); 20 C.F.R. § 655.90(b)(1)(A). U.S. workers are not considered to be "available" for purposes of the H-2A guest worker program if they insist on higher wages and better job conditions than required by U.S. conditions. *See Hernandez Flecha*, 567 F.2d at 1156-57 ("A person who is willing only if certain conditions are met is not 'willing and available.'").

24. *See* U.S. Dep't of Labor, *supra* note 4; *see, e.g.*, *Elton Orchards*, 508 F.2d at 500 (holding that a business preference to hire more skilled or experienced workers is an insufficient qualification); Bernett v. Hepburn Orchards, Inc., No. JH-84-991, 1987 WL 16939, at *4-6 (D. Md. Apr. 14, 1987) (stating that a U.S. worker only needs to be "minimally qualified" for a job and need not pass an employer's subjective "ladder test").

25. 8 U.S.C. § 1188(c)(3)(A); 20 C.F.R. § 655.102(c).

workers in the United States similarly employed."[26] For example, if H-2A workers were permitted to work for lower wages than those of U.S. workers, the wages of U.S. workers would have to decrease in order to compete—an adverse effect that would be contrary to the stated purpose of the system: the protection of U.S. workers.

IV. CERTIFICATION PROCEDURE

To ensure that the two certification conditions are met, agricultural employers seeking to hire foreign guest workers must file an application and job offer with the Regional Administrator of the United States Department of Labor's Employment and Training Administration.[27] The employer must also provide a "positive recruitment plan" that the employer will use to recruit U.S. workers.[28] To be certified, the employer must comply with the application and job offer requirements and sufficiently assure that no U.S. workers were unfairly rejected or adversely affected.

A. The Application

1. Timing

Agricultural employers must file an H-2A application with the Regional Administrator "no less than 45 calendar days before the first date on which the employer estimates that the workers are needed."[29] The Regional Administrator must notify the employer in writing of any deficiencies in the agricultural employer's filed application within seven days of its receipt.[30] The employer then has five days to correct the deficiencies and file an amended application.[31] If the application is sufficient, the Regional Administrator must issue the final certification decision "no later than 20 calendar days before the date on which the workers are needed."[32]

---

26. 8 U.S.C. § 1188(a)(1)(B).

27. *See* 20 C.F.R. § 655.101(a)(1), (b)(1).

28. *Id.* § 655.102(d).

29. *Id.* § 655.101(c)(1). The timeliness of filing an H-2A application is an important aspect of the certification process. Because agricultural services need to be performed on a seasonal basis, the crops may go bad and money may be lost if the services are not provided at the appropriate time.

30. 8 U.S.C. § 1188(c)(2)(A); 20 C.F.R. § 655.101(c)(2).

31. 20 C.F.R. § 655.101(c)(2).

32. *Id. But see* 8 U.S.C. § 1188(c)(3) (stating: "The Secretary of Labor shall make, not later than *30 days* before the date such labor or services are first required to be performed, the certification . . . .") (emphasis added).

2. Persons Who May Submit an Application

Either an agricultural employer, an agent of an agricultural employer, or an association may submit an H-2A application to the Department of Labor.[33] If an application is completed by an agent of an agricultural employer, the employer must include a signed statement with the application that explicitly "authorizes the agent to act on the employer's behalf."[34] The statement can also authorize the agent to conduct interviews and hire employees on the employer's behalf.[35] However, it must expressly state "that the employer assumes full responsibility for the accuracy of the application, for all representations made by the agent on the employer's behalf, and for compliance with all regulatory and other legal requirements."[36]

If an application is filed by an association, the specific employer must be a member.[37] The association must designate on the application "whether it is: (i) [t]he sole employer; (ii) a joint employer with its employer-member employers; or (iii) the agent of its employer-members."[38] Additionally, the association must both submit documentation to allow the Regional Administrator to confirm the status of the association and provide a list of the employer-members who desire to employ H-2A guest workers.[39] If the association will be the joint or sole employer of the H-2A workers, the workers can be transferred among the association's employer-members.[40] As a joint employer, the association is subject to liability if its employer-members violate the H-2A regulations.[41]

3. Items That Must Be Submitted

The H-2A application must (1) be submitted on a form supplied by the Employment and Training Administration, (2) be accompanied by a job offer, (3) contain an agreement stating an intent to abide by federal regulations, and (4) "state the total number of [H-2A] workers the employer anticipates employing . . . during the covered period of employment."[42] The application must also provide sufficient information to allow the Regional Administrator to determine whether the services needed are "'of a temporary or seasonal

---

33. 20 C.F.R. § 655.101(a)(1)-(3).
34. *Id.* § 655.101(a)(2).
35. *Id.*
36. *Id.*
37. *See id.* § 655.101(a)(3).
38. *Id.*
39. *Id.*
40. *See* 8 U.S.C. § 1188(c)(3)(B)(iv) (2000).
41. *See, e.g.*, Salazar-Calderon v. Presidio Valley Farmers Ass'n, 765 F.2d 1334, 1345-46 (5th Cir. 1985).
42. 20 C.F.R. § 655.101(b).

nature.'"[43]

Prior to the Regional Administrator's final certification decision, the employer can amend the application "to increase the number of [H-2A] workers requested."[44] If the increase is less than twenty percent of the total number of H-2A workers originally requested, the employer is not required to spend additional time recruiting U.S. workers.[45] However, if the increase is greater than twenty percent, the Regional Administrator can order an additional recruitment period.[46] The Regional Administrator may waive the additional recruitment period if the increased need "could not have been foreseen" and the employer's "crops or commodities will be in jeopardy prior to the expiration of an additional recruitment period."[47]

B. Job Offer

A "job offer," or "clearance order," sets forth the "material terms and conditions of employment" and is "'essentially an offer for a contract of employment.'"[48] Simultaneous with the filing of the application and job offer with the Regional Administrator, a copy of the application and job offer is submitted to the state's employment service agency, or the "local office."[49] The job offer is then used "to prepare a local job order and . . . to recruit U.S. workers in the area of intended employment."[50] The state's employment service agency also generates an "agricultural clearance order," which is used "to recruit [U.S.] workers in other geographical areas [once] the employer's H-2A application is accepted for consideration and the [job offer] is approved by the [Regional Administrator]."[51]

---

43. *Id.* § 655.101(g).

44. *Id.* § 655.101(d). Increasing the number of requested H-2A workers is not considered a "'minor technical amendment[].'" *Id.* § 655.101(e). An employer may request "minor technical amendments" to the application at any time prior to the Regional Administrator's certification determination. *Id.* The Regional Administrator will approve minor amendments if the changes "are justified and will have no significant effect upon the RA's ability to make the labor certification determination." *Id.*

45. *Id.* § 655.101(d). If the employer originally sought to hire less than ten H-2A workers, the increase cannot be more than fifty percent, rather than the standard twenty percent. *Id.*

46. *See id.*

47. *Id.*

48. *Id.* § 655.100(b); Bernett v. Hepburn Orchards, Inc., No. JH-84-991, 1987 WL 16939, at *2 (D. Md. April 14, 1987) (quoting W. Colo. Fruit Growers Ass'n v. Marshall, 473 F. Supp. 693, 696 (D. Colo. 1979)).

49. 20 C.F.R. § 655.101(c)(4).

50. *Id.*

51. *Id.*

1. Job Offer Content Requirements

The job offer must extend "the same benefits, wages, and working conditions [to U.S. workers] which the employer is offering, intends to offer, or will provide to H-2A workers."[52] It may not "impose on U.S. workers any restrictions or obligations which will not be imposed on . . . H-2A workers."[53]

Federal regulations require all job offers to provide "minimum benefit[s], wage[s], and working condition[s]" in an effort "to protect similarly employed U.S. workers from [the] adverse effect[s]" that may result from the employment of H-2A guest workers.[54] To ensure this protection, employers must include the following terms in job offers:

> (1) that if workers "are not reasonably able to return to their residence within the same day," the agricultural employer will provide them with housing at no cost;[55]
> (2) that the employer will offer housing for workers' families, if "it is the prevailing practice in the area and occupation of intended employment[;]"[56]
> (3) that the employer will provide workers with workers' compensation insurance at no cost to the worker;[57]
> (4) that the employer will furnish the workers at no cost with the tools, supplies, and equipment needed to perform the required agricultural services, unless "the common practice in the particular area . . . and occupation [is] for workers to provide [their own] tools and equipment[;]"[58]
> (5) that when "centralized cooking and eating facilities" are not available, the employer will provide workers either with cooking facilities or three meals a day at a clearly stated low cost (not to be more than $5.26 per day);[59]
> (6) that once a worker fulfills fifty percent of the work contract period, the employer will reimburse "the worker for costs incurred . . . for transportation and daily subsistence" from the worker's permanent home to the U.S. job site;[60]

---

52. *Id.* § 655.102(a).

53. *Id.*

54. *Id.* § 655.102(b).

55. Id. § 655.102(b)(1).

56. U.S.C. § 1188(c)(4) (2000). The housing, which may be rental or public, must meet either the local, state, or federal standards. *Id.* Even if the employer is only providing housing for the worker, that housing must meet either the applicable local, state, or federal standard. *See id.*

57. 20 C.F.R. § 655.102(b)(2).

58. *Id.* § 655.102(b)(3). If a worker is provided with tools and fails to return them at the end of employment, "the employer may charge the worker for reasonable costs" of the tools. *Id.*

59. *Id.* § 655.102(b)(4).

60. *Id.* § 655.102(b)(5)(i). The amount to be reimbursed for transportation "shall be no less (and shall not be required to be more) than the most economical or reasonable similar common carrier transportation charges for the distances involved." *Id.*

(7) that once a worker fulfills the entire work contract period, the employer will "provide or pay for the worker's transportation and daily subsistence from the place of employment" to the worker's permanent home;[61]
(***) that workers are guaranteed the opportunity to work "for at least three-fourths of the work days" in the work contract period,[62] unless a worker loses this "three-fourths guarantee" by being terminated for cause,[63] abandoning the position,[64] or being displaced by an available U.S. worker;[65]
(8) that the employer will keep accurate records of the daily number of hours that each worker performs, the nature of the work completed, the number of work hours offered daily to each worker, and "the worker's earnings per pay period[;]"[66]
(9) that the employer will provide the worker with a statement "on or before each payday" stating the "total earnings for the pay period," the "hourly rate . . . of pay," the number of "hours of employment . . . offered to the worker," the number of "hours actually worked by the worker,"[67] and "[a]n itemization of all deductions made from the worker's wages[;]"[68]
(10) that the employer will supply workers with a schedule of when they will be paid, which is either "in accordance with the prevailing practice in the area of intended employment, or at least twice monthly[,] whichever is more frequent[;]"[69] and
(11) that the employer will provide workers with a clear statement of "all deductions not required by law" that will be taken out of their paycheck.[70]

2. Assurances Included in the Job Offer

An employer must expressly promise in the job offer to abide by the "assurances" set forth in the federal regulations.[71] By doing so, the employer guarantees the following: (1) that the employer is not seeking H-2A workers

---

61. *Id.* § 655.102(b)(5)(ii).

62. *Id.* § 655.102(b)(6)(i). If a worker refuses to work a day at no fault of the employer, that day still counts towards the "three fourths guarantee" work period. *Id.* § 655.102(b)(6)(iii). If an employer fails to furnish workers with employment for at least three-fourths of the work contract period, the employer must "pay [the] worker[s] the amount which [they] would have earned had [they] . . . worked for the guaranteed number of days." *Id.* § 655.102(b)(6)(i).

63. *Id.* § 655.102(b)(11).

64. *Id.*

65. *Id.* § 655.102(b)(6)(iv).

66. *Id.* § 655.102(b)(7)(i).

67. *Id.* § 655.102(b)(8)(i)-(iv).

68. *Id.* § 655.102(b)(8)(v).

69. *Id.* § 655.102(b)(10).

70. *Id.* § 655.102(b)(13). "All deductions shall be reasonable and, if the employer is subject to the Fair Labor Standards Act, must not cause the worker's pay rate to dip below the federal minimum wage rate." *Id.*; *see* Arriaga v. Fla. Pac. Farms, L.L.C., 305 F.3d 1228, 1236 & n.9 (11th Cir. 2002).

71. *See* 20 C.F.R. § 655.103.

because of a labor dispute;[72] (2) that the employer will adhere to all "applicable federal, State, and local employment-related laws an regulations, including employment-related health and safety laws;"[73] (3) that "[n]o U.S. worker will be rejected for or terminated from employment for other than a lawful job-related reason" and that the local employment service agency will be notified of all such rejections and terminations;[74] (4) that the employer will independently recruit U.S. workers for the positions "until the foreign workers have departed for the employer's place of employment[;]"[75] and (5) that the employer will hire "any qualified, eligible U.S. worker who applies to the employer" after the H-2A work contract period has begun, if at least fifty percent of the work contract period remains.[76]

3. Adverse Effect Wage Rate

Each job offer must include a promise to "pay the workers at least the adverse effect wage rate [(the "AEWR")] in effect at the time the work is performed, the prevailing hourly wage rate, or the legal federal or State minimum wage rate, whichever is highest."[77] The AEWR is the "annual weighted average hourly wage rate for field and livestock workers (combined) for [a specific] region as published annually by the U.S. Department of Agricultural."[78] The AEWR, which must be published for each state "at least once in each calendar year,"[79] should never drop below the federal minimum wage rate.[80] If the prevailing wage rate in a region is higher than the published AEWR for the region, the agricultural employer must pay workers the prevailing wage rate.[81]

---

72. *Id.* § 655.103(a).

73. *Id.* § 655.103(b).

74. *Id.* § 655.103(c).

75. *Id.* § 655.103(d).

76. *Id.* § 655.103(e). This assurance is known as the "fifty-percent rule." *Id.* "The goal behind the rule is clear—to provide work for qualified domestic laborers." Clarke v. Gardenhour Orchards, Inc, No. JH-84-4419, 1987 WL 48234, at *3 (D. Md. July 23, 1987). If qualified U.S. workers are hired under this rule, the H-2A workers' employment will be terminated. *See* 20 C.F.R. § 655.102(b)(6)(iv).

77. *Id.* § 655.102(b)(9)(i).

78. *Id.* § 655.107(a). The Agricultural Job Opportunities, Benefits, and Security Act of 2005, currently under consideration by Congress, would change the way that the AEWR is set. *See* Agricultural Job Opportunities, Benefits, and Security Act of 2005, S. 359, 109th Cong. (2005).

79. 20 C.F.R. § 655.107(a).

80. *Id.* § 655.107(c).

81. *Id.* § 655.107(b).

C. The Regional Administrator's Determination

When the Regional Administrators receive the agricultural employers' applications and job offers, they "promptly review the application[s]" to determine "if there are deficiencies which render the application not acceptable for *consideration*."[82] To make this determination, Regional Administrators review the applications to assure that they are in compliance with the "timeliness" requirements set forth in the federal regulations.[83] They also consider whether the "benefits, wages, and working conditions" in the job offer will adversely affect similarly situated U.S. workers.[84] If the Regional Administrator determines that the "temporary alien agricultural labor certification application meets the adverse effect criteria necessary for processing," then the application is approved for *consideration*,[85] and the availability of U.S. workers is tested.[86]

Regional Administrators will reject applications and job offers for *consideration* if they are submitted during a lockout or strike[87] or there is "[in]sufficient time to test the availability of U.S. workers."[88] If a Regional Administrator rejects an application for consideration, written notice must be sent to the employer "within seven calendar days of the date the application was received by the [Regional Administrator]."[89] The notice must state the following: (1) the grounds for denial, "citing the relevant regulatory standards[;]"[90] (2) the employer's option to resubmit a "modified application" within five days of receiving the notice of denial;[91] and (3) the employer's right "to request an expedited administrative review or a *de novo* administrative hearing before an administrative law judge of the nonacceptance."[92] The notice shall explain that, if the employer chooses to appeal the denial of consideration, the employer must file a written request with the Chief Administrative Law Judge of the Department of Labor within seven days of receiving notice of the denial.[93]

If an agricultural employer requests an expedited administrative review of

---

82. *Id.* § 655.100(a)(1) (emphasis added). When Regional Administrators approve applications for consideration, they are not making final determinations. *See id.* § 655.100(b). By accepting an application for consideration, the Regional Administrator only certifies that the pending application "meets the adverse effect criteria necessary for processing." *Id.*

83. *Id.* § 655.104(b).

84. *See id.*

85. *Id.* § 655.100(b).

86. *See id.* § 655.104(b).

87. *Id.* § 655.103(a).

88. *Id.* § 655.100(a)(3).

89. *Id.* § 655.104(c).

90. *Id.* § 655.104(c)(1).

91. *Id.* § 655.104(c)(2).

92. *Id.* § 655.104(c)(3).

93. *Id.*

the Regional Administrator's denial of consideration, the Regional Administrator must send the employer's application file to the Chief Administrative Law Judge "by means normally assuring next-day delivery."[94] The employer's case is then "immediately assign[ed to] an administrative law judge[, who] review[s] the record [of the Regional Administrator's denial] for legal sufficiency."[95] During this review, the administrative law judge can neither remand the case to a lower court nor gather any additional evidence.[96] The judge must decide whether to "affirm, reverse, or modify the [Regional Administrator]'s denial" of consideration "[w]ithin five working days after receipt of the case file."[97] The employer must then be immediately notified by "next-day delivery" of the judge's decision and reasoning.[98] The decision of the administrative law judge is final and will not receive further review from the Department of Labor.[99]

If an agricultural employer requests a de novo administrative hearing on the Regional Administrator's denial of consideration instead of an expedited administrative review, the process only differs in that an administrative law judge conducts a full hearing, after which the judge has ten days to "either affirm, reverse, or modify the [Regional Administrator]'s determination."[100] If so requested by the employer, the administrative law judge must conduct the hearing "within five working days after the . . . receipt of the case file."[101]

D. Recruitment of U.S. Workers

An agricultural employer must submit a "positive recruitment plan" directed towards U.S. workers to the Regional Administrator in addition to the H-2A application and job offer.[102] The recruitment plan must conform with the following requirements: (1) it must be in writing; (2) it must "include a description of [the employer's] recruitment efforts (if any) made prior to the actual submittal of the application[;]" and (3) it must state "how the employer will engage in [the] positive recruitment of U.S. workers" after the H-2A application has been filed.[103] An agricultural employer's recruitment plan for U.S. workers must expend "the same level of effort [in recruiting] as non-H-2A agricultural employers" of similar size in the same geographic area.[104]

If the Regional Administrator accepts the agricultural employer's

---

94. *Id.* § 655.112(a)(1).
95. *Id.*
96. *Id.*
97. *Id.* § 655.112(a)(2).
98. *Id.*
99. *Id.*
100. *Id.* § 655.112(b)(1)(iii)-(2).
101. *Id.* § 655.112(b)(1)(ii).
102. *Id.* § 655.102(d).
103. *Id.*
104. *Id.*

application for consideration, the Regional Administrator must notify the local employment service agency of the acceptance "no later than seven calendar days after the application was received."[105] When the local employment service agency receives the H-2A application, it "promptly prepare[s] an agricultural clearance order which will permit the recruitment of U.S. workers . . . on an intrastate and interstate basis."[106]

The Regional Administrator determines which states are "potential sources of [available] U.S. workers" for recruitment purposes and submits the job offer portion of the H-2A application to those states.[107] The agricultural employer is required to positively and actively participate in the recruitment of able and willing U.S. workers until the H-2A guest workers leave their home for the place of employment.[108] This participation includes placing job advertisements in newspapers and on radio commercials in the areas deemed appropriate by the Regional Administrator.[109] All advertisements placed by the employer must disclose the following: (1) "the nature and anticipated duration" of the work; (2) a payrate no lower than the AEWR; (3) that the employer will provide housing and work tools and reimburse transportation expenses; and (4) that interested potential candidates should apply at their local employment service office.[110] Under the required participation, agricultural employers must also "contact[] farm labor contractors, migrant workers and other potential workers in other areas of the State and/or Nation," along with any organizations that may help in the recruitment of U.S. workers.[111]

To determine whether the agricultural employer has sufficiently attempted to recruit U.S. workers, the Regional Administrator compares "the normal recruitment practices of non-H-2A agricultural employers in the [geographic] area and the kind and degree of recruitment efforts which the potential H-2A employer made to obtain H-2A workers."[112] The employer's efforts to recruit U.S. workers must at least match both the recruitment practices of non-H-2A agricultural employers of similar size in the same area and the employer's own efforts to recruit H-2A workers.[113] The Regional Administrator must conclude whether the employer has satisfied the recruitment requirements no less than

---

105. *Id.* § 655.104(a).

106. *Id.*

107. *Id.* § 655.105(a). Federal regulations require that the job offer be submitted to states via an "intrastate" and "interstate" clearance system. *Id.* This clearance system was established by the Wagner-Peyser Act, 29 U.S.C. §§ 49-49*l*-2 (2000), and its primary purpose is to help unemployed U.S. workers find employment. Frederick County Fruit Growers' Ass'n v. Marshall, 436 F. Supp. 218, 220 (W.D. Va. 1977).

108. 20 C.F.R. § 655.103(d).

109. *Id.* § 655.103(d)(2).

110. *Id.* § 655.103(d)(2)(i)-(ii).

111. *Id.* § 655.103(d)(3)-(4).

112. *Id.* § 655.105(a).

113. *Id.*

"20 calendar days before the date of need [for services] specified in the application."[114] If the Regional Administrator finds that the employer's efforts to recruit U.S. workers were insufficient, the Regional Administrator must promptly notify the employer that "the temporary alien agricultural labor certification" to hire H-2A workers has been denied.[115] The employer then has the right to an administrative review or a de novo administrative hearing on the denial of *certification*.[116]

Even if an agricultural employer is certified to hire H-2A workers, the employer must hire any qualified U.S. workers who apply for a job within the first fifty percent of the H-2A guest workers' work contract period, resulting in the displacement of the H-2A workers.[117] This requirement, although harsh, furthers the purpose of the H-2A application and certification procedure: the hiring and protection of U.S. workers whenever possible.

V. PROBLEMS WITH THE CURRENT CERTIFICATION PROCEDURES

The current H-2A certification system is deficient and needs to be amended. U.S. workers are not receiving sufficient protection. Although studies indicate that there is no shortage of U.S. agricultural workers, the number of agricultural employers certified to hire H-2A guest workers continues to increase.[118]

Hoping that the deficiencies will escape the notice of the Regional Administrator, agricultural employers who prefer to hire H-2A workers over U.S. workers may intentionally fail to satisfy the H-2A guest worker program's regulatory requirements in an attempt to deter U.S. workers from applying for positions. One effective method of deterrence is to submit a job offer that contains wages and working conditions well below the prevailing conditions in the geographic area. For example, in a letter written to a Regional Administrator by a U.S. agricultural worker advocate, the advocate references a deficient job offer that the Regional Administrator had approved for consideration.[119] The advocate reveals that the wage rates and working conditions contained in the job offer were below those prevailing and offered

---

114. *Id.* § 655.105(d).

115. *Id.*

116. *Id.* § 655.105(e).

117. 8 U.S.C. § 1188 (c)(3)(B)(i) (2000); 20 C.F.R. § 655.103(e); *see id.* § 655.102(b)(6)(iv).

118. *See* Michael Holley, *Disadvantaged by Design: How the Law Inhibits Agricultural Guest Workers from Enforcing Their Rights*, 18 HOFSTRA LAB. & EMP. L.J. 575, 576 (2001) (noting that "the use of H-2A workers . . . tripled from 1995 to 1999," although "a 1997 study by the U.S. General Accounting Office . . . show[ed] that there [was] no shortage of farm workers in the United States").

119. Letter from Gregory S. Schell, Attorney, to Toussaint Hayes, Reg'l Adm'r, Employment & Training Admin, U.S. Dep't of Labor 1-2 (Jan. 19, 1998) [hereinafter Jan. 19, 1998 Letter] (on file with the Tennessee Law Review).

by similar non-H-2A agricultural employers in the same geographic area.[120] In a separate letter to the Regional Administrator, the advocate points out that the H-2A employer's job offer, which the Regional Administrator had accepted for consideration, offered to pay only "40 [cents] per 5/8 bushel bag," while "the prevailing wage rate . . . [in the area] was 1.8 [cents] per pound."[121] The advocate correctly informs the Regional Administrator that the H-2A agricultural employer had failed to "demonstrat[e] that the earnings under the 40 [cents] per bucket rate [were] equal to or exceed[ed] the average hourly earnings of workers paid by the pound at the prevailing rate."[122]

Another key method of deterrence practiced by employers is to require certain job qualifications in the job offer that are not required of similar agricultural workers in the same geographic area. For example, in a letter written to a Regional Administrator, a U.S. agricultural worker advocate opposes the Regional Administrator's approval for consideration of an H-2A job offer that required "[w]orker[s] to produce enough units under the piece-rate system to earn not less than the current federal minimum hourly wage."[123] The offer implied that, if workers could not meet this "production standard," they would be "subject to termination."[124] After noting that federal regulations only permit agricultural employers to impose job qualifications if the prevailing practice in the occupation is to do so,[125] the advocate informs the Regional Administrator of an authoritative state survey which "found that the prevailing practice among [that industry in the state was to] not impose productivity standards as job qualifications."[126]

Besides intentional efforts to deter U.S. workers, there have also been deficiencies in approved H-2A agricultural job offers regarding the recruitment of U.S. workers. For example, in a letter addressed to a Regional Administrator, an agricultural worker advocate argues that the Regional Administrator erroneously approved an H-2A application and job offer for consideration because the employer had failed to include a positive

---

120. *See id.* at 2-12. Non-H-2A agricultural employers in the region offered a "piece-rate wage" (the worker is paid by the number of "pieces" picked, for example, twenty cents per bucket of apples), while the H-2A agricultural employer offered an hourly rate. *See id.* at 6-7. The hourly rate wage would be substantially less than what the workers would earn on a piece-rate wage. *See id.* at 7.

121. Letter from Gregory S. Schell, Attorney, to Toussaint Hayes, Reg'l Adm'r, Employment & Training Admin, U.S. Dep't of Labor 2 (Feb. 23, 1998) (on file with the Tennessee Law Review).

122. *Id.*

123. Letter from Gregory S. Schell, Attorney, to Daniel L. Lowry,Reg'l Adm'r, Employment & Training Admin, U.S. Dep't of Labor 1-2, 8(Apr. 29, 1991) [hereinafter Apr. 29, 1991 Letter] (on file with the Tennessee Law Review).

124. *Id.* at 2.

125. *Id.* (citing 20 C.F.R. § 655.102(c)); *see supra* text accompanying notes 23-26.

126. Apr. 29, 1991 Letter, *supra* note 123, at 2.

Administrator.[132] The H-2A certification procedure is heavily weighted in favor of the employers, despite the fact that the purpose of the process is to protect U.S. workers.

Granting U.S. workers and their advocates the right to challenge an agricultural employer's pending H-2A application, which has been approved for consideration, would improve, although not entirely remedy, the problems with the current H-2A system. This right would advance the Employment and Training Administration's watchdog function and force employers to be more conscientious in their efforts to satisfy the H-2A requirements, for fear of challenge by adversely affected U.S. workers. Moreover, with a system of review for pending H-2A applications, Regional Administrators would be more vigilant with employers to assure that the employers had met all federal regulations and that no U.S. workers were "able, willing, qualified, and . . . available" to fill the position.

To ensure these results, a federal regulation should be implemented that would permit U.S. workers or their advocates to appeal a Regional Administrator's decision to consider an agricultural employer's H-2A application. Their right to appeal should mirror the rights afforded to employers when H-2A applications are denied for consideration.

In challenging a Regional Administrator's approval for *consideration* of an application, a U.S. worker should have the right to appeal the decision to an administrative law judge. I propose the following process: First, the worker (or workers) wishing to appeal the decision would submit a copy of the H-2A job offer and any evidence of existing deficiencies in the H-2A application or job offer to the administrative law judge. Second, the Regional Administrator would have to submit the H-2A agricultural employer's file and the H-2A application to the administrative law judge. If the administrative law judge finds merit in the challenge, then the Regional Administrator's decision to consider the application for certification would be reversed. The H-2A employer, the Regional Administrator, and the worker challenging the determination would all receive notice of the administrative law judge's decision, and the employer would then be permitted to amend the H-2A application or job offer and resubmit it once the deficiencies were remedied.

Overall, this appellate process would be similar to the rights that currently exist under the H-2A certification system for agricultural employers. Employers have the right to appeal unfavorable decisions. Why should adversely affected U.S. workers not have the same rights? Critics of appellate rights for U.S. worker have answered this question in two ways.

First, some critics may argue that allowing U.S. workers and their advocates to challenge the decisions of the Regional Administrator would increase the delay in the already drawn out H-2A certification process. But these critics ignore the fact that the proposed appellate processes for U.S.

---

132. *See supra* text accompanying notes 94-101.

workers would have strict time requirements similar to those imposed on the employers' appeal process.[133] Therefore, the addition of a formal avenue for U.S. workers to challenge a Regional Administrator's determination would not create any additional delay.

A second argument that critics may advance is that a formal avenue of appeal for U.S. workers would be a waste of state and federal resources because U.S. workers would bring meritless challenges to harass agricultural employers and delay the process. Although this is a valid concern, this criticism fails to negate the fact that U.S. workers deserve a right to challenge Regional Administrators' decisions—a right that H-2A employers currently enjoy. Without such a right for U.S. workers, those meant to be protected by the certification system would be prevented from seeking to have that protection enforced.

VII. CONCLUSION

The current certification process is in need of change. The purpose of the process is to protect U.S. workers, but the reality is that U.S. workers are not being protected. When able and willing U.S. workers are available to fill positions, agricultural employers are required by law to hire them. Under the current system, however, agricultural employers, despite the availability of U.S. workers, are intentionally and unintentionally frustrating the program's purpose, resulting in the increased employment of foreign guest workers at the expense of U.S. workers. The U.S. Department of Labor has failed in its enforcement functions to prevent this from happening.

The current certification system should be amended to allow U.S. workers and their advocates to challenge agricultural employers' pending H-2A applications. This formal avenue of appeal would better protect U.S. workers and more effectively fulfill the purpose of the certification process: the protection of U.S. workers. Agricultural employers are continually discovering cost advantages to hiring foreign guest workers under the H-2A system. A formal avenue of appeal for U.S. workers and their advocates would create more accountability where little currently exists and result in the enforcement of the H-2A safeguards.

LESLIE GREEN

133. *See supra* text accompanying notes 94-101.

CIVIL RIGHTS—ADEA—THE AVAILABILITY OF DISPARATE IMPACT CLAIMS

Smith v. City of Jackson, 544 U.S. 228 (2005).

I. INTRODUCTION

Police and other public safety officers sued the city of Jackson, Mississippi for violating the Age Discrimination in Employment Act of 1967 (the "ADEA").[1] The defendant granted raises to all city employees in an effort to "bring the starting salaries of police officers up to the regional average."[2] Those officers who had served less than five years with the police force, however, received proportionately larger raises than those officers with more experience.[3] Plaintiffs claimed that the disproportionate raises for those officers with more seniority violated the ADEA[4] because most officers aged 40 and over had more than five years experience.[5] The plaintiffs alleged that the defendant, in violation of the ADEA, "deliberately discriminated against them because of their age ([a] 'disparate treatment' claim) and that they were 'adversely affected' by the [salary] plan because of their age ([a] 'disparate impact' claim)."[6] The United States District Court for the Southern District of Mississippi granted the defendant summary judgment on both claims.[7] The United States Court of Appeals for the Fifth Circuit, finding that "disparate-impact claims are categorically unavailable under the ADEA," affirmed the District Court's dismissal of the plaintiffs' disparate impact claim.[8] On certiorari to the United States Supreme Court, *held*, affirmed.[9] The Court, however, did hold that disparate impact recovery is cognizable under the

---

1. Smith v. City of Jackson, 544 U.S. 228, 230 (2005).

2. *Id.*

3. Petitioners contended that the increases were "less generous to officers over the age of 40 than to younger officers." *Id.* The ADEA protects "individuals 40 and over." 29 C.F.R. § 1625.2(a) (2004).

4. Discrimination in Employment Act of 1967, 29 U.S.C. §§ 621-634 (2000).

5. *Smith*, 544 U.S. at 231.

6. *Id.*; *see also infra* text accompanying notes 11-13 for an explanation of disparate treatment and disparate impact claims.

7. *Smith*, 544 U.S. at 231.

8. *Id.* at 231. The Court of Appeals also held that the Plaintiffs were entitled to more discovery on the issue of intent with respect to the disparate treatment ruling. *Id.*

9. *Id.* at 232.

ADEA.[10] *Smith v. City of Jackson*, 544 U.S. 228(2005).

II. DISPARATE IMPACT AS A COGNIZABLE CLAIM UNDER THE ADEA

Plaintiffs may want to bring both disparate treatment and disparate impact claims for discrimination under the ADEA.[11] While a successful disparate treatment claim requires showing that the employer intended to discriminate,[12] a disparate impact claim does not require proof of intentional discrimination; a facially neutral but discriminatory practice may form the basis for a disparate impact claim.[13] While lower federal courts have been unable to come to a consensus of whether to permit disparate impact claims under the ADEA,[14] the Supreme Court has shown a willingness to permit disparate impact claims under some civil rights legislation, such as Title VII of the Civil Rights Act of 1964 ("Title VII").[15] Until *Smith*, however, the Court had not decided whether a disparate impact claim is permissible under the ADEA.[16] In *Smith*, the Court held that a disparate impact claim is cognizable under the ADEA.[17]

A. Development of the Law to Protect Against Age Discrimination

During the Civil Rights Movement in the 1960s, Congress passed civil rights legislation, such as Title VII and ADEA, to help protect the civil rights of minorities and other protected groups.[18] "Congress chose not to include age within discrimination forbidden by Title VII of the Civil Rights Act . . . being

---

10. *Id.* While the Court affirmed the judgment, a majority of justices disagreed with the Court of Appeals' holding that the ADEA does not recognize disparate impact claims. *Id.*

11. *See id.* at 231.

12. *See id.*; BARBARA LINDEMANN SCHLEI & PAUL GROSSMAN, EMPLOYMENT DISCRIMINATION LAW, 13-15 (2d ed. 1983) (noting that a defendant may successful defend against a discrimination suit by giving a "legitimate, nondiscriminatory reason for its actions" that the plaintiff cannot prove is "a pretext to mask an illegal motive.").

13. *See* Wards Cove Packing Co. v. Atonio, 490 U.S. 642, 645-46 (1989).

14. Zachary L. Karmen, Annotation, *Disparate Impact Claims Under Age Discrimination Act of 1967*, 186 A.L.R. FED. 1, 1 (2003) ("The Circuit courts . . . are divided on the viability of a disparate impact claim under the ADEA.").

15. *See* Griggs v. Duke Power Co., 401 U.S. 424, 430 (1971) ("Under [Title VII], practices procedures, or tests neutral on their face, and even neutral in terms of intent, cannot be maintained if they operate to 'freeze' the status quo of prior discriminatory employment practices."). While disparate impact availability under the ADEA has been disputed, it is undisputed among the federal courts that the ADEA protects plaintiffs from negative age-based disparate treatment. Karmen, *supra* note 14, at 1.

16. *Smith*, 544 U.S. at 231 ("Despite the age of the ADEA, it is a question that we have not yet addressed.").

17. *Id.* at 236.

18. *See, e.g.*, Age Discrimination and Employment Act of 1967, 29 U.S.C. §§ 621-34 (2000); Title VII of the Civil Rights Act of 1964, 42 U.S.C. §§ 2000e to 2000e-17 (2000).

aware that there were legitimate reasons as well as invidious ones for making employment decisions on age."[19] Instead of including age as a category in Title VII, Congress asked the Secretary of Labor to study the issue.[20]

That study, known as the Wirtz Report, "is a major element in the legislative history" of the ADEA.[21] Wirtz concluded that "age discrimination was a serious problem, but one different in kind from discrimination on account of race."[22] The report briefly discusses disparate impact claims, stating that "'[a]ny formal employment standard which requires, for example, a high school diploma will obviously work against the employment of many older workers—unfairly if, despite his limited schooling, an older worker's years of experience have given him the relevant equivalent of a high school education.'"[23] After the Wirtz Report was submitted to Congress, Congress asked the Secretary of Labor for proposed legislation and held hearings which "dwelled on unjustified assumptions about the effect of age on ability to work."[24] In considering the ADEA, members of Congress focused on older workers' difficulties in obtaining jobs and job interviews.[25] Congress, after deliberations, passed the ADEA in 1967.[26]

The ADEA and Title VII share language and construction.[27] When describing the two statutes in 1978, Justice Marshall wrote: "There are important similarities between the two statutes, to be sure, both in their aims—the elimination of discrimination from the workplace—and in their substantive prohibitions. In fact, the prohibitions of the ADEA were derived *in haec verba* from Title VII."[28] Because of these similarities, many guidelines for interpreting the ADEA come from case law interpreting Title VII.[29] One

---

19. Gen. Dynamics Land Sys., Inc. v. Cline, 540 U.S. 581, 586-87 (2004) (citation omitted).

20. *Id.* at 587.

21. Michael Evan Gold, *Disparate Impact Under the Age Discrimination in Employment Act of 1967*, 25 BERKELEY J. EMP. & LAB. L. 1, 6 (2004) (discussing the legislative history of the ADEA).

22. *General Dynamics*, 540 U.S. at 587.

23. *Smith*, 544 U.S. at 235 n.5 (Stevens, J., plurality opinion) (quoting Report of the Secretary of Labor, The Older American Worker: Age Discrimination in Employment 21 (June 1965), *reprinted in* U.S. Equal Employment Opportunity Commission, Legislative History of the Age Discrimination in Employment Act (1981) [hereinafter Wirtz Report]).

24. *General Dynamics*, 540 U.S. at 587-88.

25. *See id.* at 588. In *General Dynamics*, the Court noted that Congress's concerns centered around providing opportunities for employment to the thousands of workers 45 and over. *Id.* at 588-89.

26. Age Discrimination and Employment Act of 1967, Pub. L. No. 90-202, 81 Stat. 602 (codified as amended in 29 U.S.C. §§ 621-34 (2000)).

27. *Smith*, 544 U.S. at 234 (Stevens, J., plurality opinion) (citing Lorillard, Inc. v. Pons, 434 U.S. 575, 584 (1978)).

28. *Lorillard*, 434 U.S. at 584.

29. Nathan E. Holmes, Comment, *The Age Discrimination in Employment Act of 1967:*

such case is *Griggs v. Duke Power Co.*[30]

B. The Griggs Court's Establishment of Disparate Impact Claims in Civil Rights Cases

The Supreme Court first authorized the bringing of disparate impact claims under Title VII in *Griggs*.[31] In that case, a group of African-American employees challenged their employer's high school diploma and standardized general intelligence test requirements on the grounds that those requirements were not "significantly related to successful job performance," and that they "operate[d] to disqualify Negroes at a substantially higher rate than white applicants."[32]

The Court found that although the employer's practice did not have a discriminatory motive, "good intent or absence of discriminatory intent does not redeem employment procedures or testing mechanisms that operate as 'built-in headwinds' for minority groups and are unrelated to measuring job capability." [33] The Court stated that Title VII "proscribes not only overt discrimination but also practices that are fair in form, but discriminatory in operation."[34] Prohibited practices include those practices that "cannot be shown to be related to job performance."[35] "The touchstone is business necessity."[36]

C. Application of Griggs to the ADEA

Because the Supreme Court stated in 1979 that "the ADEA and Title VII share a common purpose" and have nearly identical wording and construction, lower courts have applied the Court's rationale in *Griggs* to cases involving the ADEA.[37] But two differences between the statutes make it unclear whether

---

Are Disparate Impact Claims Available?, 69 U. CIN. L. REV. 299, 301 (2000).

30. 401 U.S. 424 (1971).

31. *See id.* at 430.

32. *Id.* at 425-26.

33. *Id.* at 432.

34. *Id.* at 431.

35. *Id.*

36. *Id.*

37. Oscar Mayer & Co. v. Evans, 441 U.S. 750, 756 (1979). *See, e.g.*, Smith v. City of Des Moines, 99 F.3d 1466, 1469-70 (8th Cir. 1996) (noting the similarities between Title VII and the ADEA and that the court had "on several occasions applied disparate impact analysis to age discrimination claims"); EEOC v. Borden's Inc., 724 F.2d 1390, 1394-95 (9th Cir. 1984) ("While it is true that the disparate impact theory first arose in cases under Title VII . . ., the similar language, structure, and purpose of Title VII and the ADEA . . . has led us to adopt disparate impact in cases under the ADEA."); Geller v. Markham, 635 F.2d 1027, 1032 (2d Cir. 1980) (noting the similarity between Title VII and the ADEA and applying the same disparate impact analysis).

the ADEA permits disparate impact claims under *Griggs*.

First, the ADEA contains the Reasonable Factors Other than Age provision ("RFOA").[38] This language narrows the ADEA because it allows employers "to take any action otherwise prohibited . . . where the differentiation is based on reasonable factors other than age."[39] In interpreting the RFOA provision, the Equal Employment Opportunity Commission (the "EEOC") stated that an employment practice that "has an adverse impact on individuals within the protected age group . . . can only be justified as a *business necessity*," if the practice is not based on age.[40]

Second, Congress amended Title VII in 1991[41] in response to the Supreme Court's holding in *Wards Cove Packing Co. v. Atonio*,[42] but it did not amend the ADEA.[43] In *Wards Cove*, the Supreme Court, in limiting disparate impact claims and narrowly reading Title VII, placed the burden of persuasion in such cases on the plaintiff.[44] The Court also reversed the Ninth Circuit's finding that a statistical disparity between workers of different races made a prima facie case of disparate impact.[45]

Congress's amendment, the Civil Rights Act of 1991,[46] overruled the Court's "allocation of the burdens of proof," and "reinstate[d] the burden-shifting rule of pre-*Wards Cove* cases."[47] A Title VII disparate impact plaintiff has, after the Civil Rights Act of 1991, "the burdens of production and persuasion as to the prima facie case, but the defendant bears the burdens of production and persuasion as to justifying the challenged practice as a business

---

38. 29 U.S.C. § 623(f)(1) (2000).

39. *Id.*

40. 29 C.F.R. § 1625.7(d) (2004) (emphasis added) (reflecting the Court's holding in *Griggs*).

41. *See* Civil Rights Act of 1991, 42 U.S.C. § 2000e-2(k) (2000).

42. 490 U.S. 642 (1989). In *Wards Cove*, the plaintiffs made a disparate impact claim under Title VII involving discrimination on the basis of race in the salmon cannery industry. *Id.* at 646-48.

43. *See* Holmes, *supra* note 29, at 304, 306.

44. *Wards Cove*, 490 U.S. at 660 ("The persuasion burden here must remain with the plaintiff, for it is he who must prove that it was 'because of such individual's race, color," etc. that he was denied a desired employment opportunity." (quoting 42 U.S.C. § 2000e-2(a) (1988)). Holmes states that "the Supreme Court attempted to lighten the burden of justification imposed upon a defendant in a Title VII disparate impact case" in *Wards Cove*. Holmes, *supra* note 29, at 304.

45. *Wards Cove*, 490 U.S. at 655 ("[W]e reverse the Court of Appeals' ruling that a comparison between the percentage of cannery workers who are nonwhite and the percentage of noncannery workers who are nonwhite makes out a prima facie case of disparate impact."). Statistical evidence can, however, be used in some cases to make out a prima facie case of disparate impact, but the "usefulness [of statistics] depends on all of the surrounding facts and circumstances." Int'l Bhd. of Teamsters v. United States, 431 U.S. 324, 339-40 (1977).

46. Civil Rights Act of 1991, Pub. L. No. 102-166, 105 Stat. 1071 (codified in scattered sections of 42 U.S.C.).

47. *See* Holmes *supra* note 29, at 34.

necessity."[48]

Despite many lower courts' decisions to apply *Griggs* to ADEA cases, the Supreme Court's holding in *Hazen Paper Co. v. Biggins*[49] further confused the state of the law.[50] The Court noted in *Hazen Paper* that the ADEA requires employers to evaluate employees based on their ability to perform their jobs and not their age,[51] and that "[w]hen the employer's decision *is* wholly motivated by factors other than age, the problem of inaccurate and stigmatizing stereotypes disappears."[52] An employer may consider years of service without taking age into consideration at all; "it is incorrect to say that a decision based on years of service is necessarily 'age based.'"[53] The *Hazen Paper* Court explicitly stated that it had "never decided whether a disparate impact theory of liability is available under the ADEA" and that it "need not do so here."[54] Despite the *Hazen Paper* Court's limited holding, some lower courts subsequently found that disparate impact claims were not cognizable under the ADEA.[55]

Because of (I) the disagreement over whether *Griggs* applies to the ADEA, (ii) the lack of consensus on the application of *Hazen Paper*, (iii) the Court's interpretation of age as being dissimilar to sex and race,[56] and (iv) the Court's reluctance to address the issue in the past, the Supreme Court was prompted to grant certiorari in the *Smith* case and decide whether disparate

---

48. *Id.*; *see* 42 U.S.C. § 2000e-2(k)(1)(A)(i)-(ii) (2000).

49. 507 U.S. 604 (1933).

50. In *Hazen Paper*, the Supreme Court held that dismissing an employee to prevent his pension benefits from vesting was not a disparate impact violation of the ADEA, because "age and years of service are analytically distinct" and the decision would "represent an accurate judgment about the employee – that he indeed is 'close to vesting.'" *Id.* at 611. The Court noted that "firing an employee to prevent his pension from vesting. . . is actionable under . . . ERISA . . . . [b]ut it would not, without more, violate the ADEA." *Id.* at 612.

51. *Id.* at 611.

52. *Id.*

53. *Id.*

54. *Id.* at 610. In *Smith*, the Court, acknowledging the confusion *Hazen Paper* created, stated that "there is nothing in our opinion in *Hazen Paper* that precludes an interpretation of the ADEA that parallels our holding in *Griggs*." *Smith*, 544 U.S. at 238 (Stevens, J., plurality opinion).

55. *See, e.g.*, Ellis v. United Airlines, Inc., 73 F.3d 999, 1001, 1008-09 (10th Cir. 1996) ("Although the [*Hazen Paper*] Court's holding was technically limited to the disparate treatment claim before it, one cannot read the opinion without receiving the strong impression that the Supreme Court is suggesting that the ADEA does not encompass a disparate impact claim."); EEOC v. Francis W. Parker Sch., 41 F.3d 1073, 1077 (7th Cir. 1994) ("[Employment] decisions based on criteria which merely tend to affect workers over the age of forty more adversely than workers under forty are not prohibited.").

56. *See* Gen. Dynamics Land Sys., Inc. v. Cline, 540 U.S. 581, 584, 597 (2004) (holding that the ADEA does not allow reverse age discrimination claims because "[t]he term 'age' employed by the ADEA is not, however, comparable to the terms 'race' or 'sex' employed by Title VII").

impact claims are cognizable under the ADEA.[57]

III. DISPARATE IMPACT UNDER THE ADEA: RATIONALE FOR APPLYING *GRIGGS* TO THE ADEA

In *Smith* the United States Supreme Court held that disparate impact claims are cognizable under the ADEA.[58] Clearing up an historically vague area of the law, the Court stated that it is possible for a plaintiff to recover from a defendant under a disparate impact theory for age-based discrimination.[59] Justice Stevens, writing an opinion that gathered a majority in sections I, II, and IV and a plurality in section III,[60] determined that disparate impact claims are cognizable under the ADEA after examining the text of the statute, the RFOA provision, and the Department of Labor and the EEOC interpretations of the ADEA.[61] Following a brief recitation of the facts and appellate history,[62] Justice Stevens began his opinion by discussing the legislative history of the ADEA.[63] He expressly noted the similarities in the language and parallels in the structure of the ADEA and Title VII.[64] He also noted that the ADEA's RFOA provision is not in Title VII; the RFOA provision "contains language that significantly narrows [the ADEA's] coverage."[65] Section III of Justice Stevens's opinion, which specifically analyzes why disparate impact claims are cognizable under the ADEA, was the most hotly contested section; it only obtained four signatures.[66]

---

57. Prior to the Court's decision in *Smith*, Holmes noted the need for a Supreme Court decision on whether a disparate impact claim was permitted under the ADEA. Holmes, *supra* note 29, at 300. In his Comment, Holmes argued that as more companies convert their defined benefit pension plans, which "determine employee benefits by considering years of service and rate of pay toward the end of an employee's career, [to] the newer cash-balance plans[, which] determine benefits based upon a percentage of the employee's yearly salary," more cases will be filed under the ADEA alleging disparate impact. *Id.* at 299-300. He stresses the need for a Supreme Court decision to prevent "different results based merely upon the federal circuit in which their suit is brought." *Id.* at 300.

58. *Smith*, 125 S. Ct. at 236.

59. *Id.* The Court held, however, that the plaintiffs in the case at bar did not provide sufficient evidence to prove a disparate impact claim. *Id.* at 232.

60. *Id.* at 229.

61. *Id.* at 233.40 (Stevens, J., plurality opinion).

62. *Id.* at 1539-40 (majority opinion).

63. *Id.* at 232.

64. *Id.* at 233-34.

65. *Id.* Justice Stevens refers to the RFOA provision as "an affirmative defense to liability" that is contained in the ADEA. *Id.* at 233 n.3.

66. *Id.* at 229 (Stevens, J., plurality opinion). The Chief Justice took no part in this opinion. *Id.* at 243. Justice Scalia agreed that disparate impact claims are cognizable under the ADEA, based solely on deference to the EEOC. *Id.* at 243 (concurring in part and concurring in the judgment). Justices O'Connor, Kennedy, and Thomas all disagreed and found "disparate impact claims are not cognizable under the ADEA." *Id.* at 248 (O'Connor, J., concurring in the

Justice Stevens looked at the text of the statute to examine the availability of disparate impact claims under the ADEA.[67] He concluded that Congress wanted the text of the ADEA and Title VII to have the same meaning because the two have almost identical language, structure, and purpose, and they were enacted close to one another in time.[68] Further, because Title VII and the ADEA are so similar, the Court's holding "in *Griggs* is . . . precedent of compelling importance" to whether disparate impact claims are cognizable under the ADEA.[69]

Justice Stevens then briefly analyzed the holding of *Griggs* and found significance when the Court "explained that Congress 'had directed the thrust of the Act to the *consequences* of employment practices, not simply the motivation.'"[70] The *Griggs* Court determined that intent was unimportant under a Title VII disparate impact claim and "squarely held that . . . Title VII did not require a showing of discriminatory intent."[71] He further noted that the Court's holding in *Griggs* was consistent with the Wirtz Report, which the legislature relied on in enacting the ADEA.[72]

After analyzing *Griggs*, Justice Stevens discussed pertinent sections of Title VII, comparable portions of the ADEA, and each section's relevance to disparate impact claims.[73] His broad proposition is that the text of both statutes "focuses on the *effects* of the action on the employee rather than the motivation for the action of the employer."[74] In disputing Justice O'Connor's assertions regarding § 4(a)(1) and § 4(a)(2) of the ADEA,[75] Justice Stevens contended that paragraph (a)(1) focuses on the employer's actions towards one particular individual, while the language in paragraph (a)(2) refers to the effect of the employer's actions on any employee.[76] He emphasized various portions of the statutory text to make his point clear: "Paragraph (a)(2) . . . makes it unlawful for an employer 'to limit . . . his *employees* in any way that [sic] would deprive or tend to deprive *any individual* of employment opportunities or otherwise adversely affect his status as an employee, because of *such individual's* age.'"[77] Justice Stevens's use of italics emphasized the

---

judgment).

67. *Id.* at 233 (Stevens, J., plurality opinion).

68. *Id.*

69. *Id.* at 234 Disparate impact claims are cognizable under Title VII. *See* Griggs v. Duke Power Co., 401 U.S. 424, 430 (1971).

70. *Smith*, 544 U.S. 234 (Stevens, J., plurality opinion) (quoting *Griggs*, 401 U.S. at 432).

71. *Id.* at 235.

72. *See id.* at 235 n.5.

73. *Id.*

74. *Id.* at 236.

75. *See id.* at 236 n.6. Sections 4(a)(1) and 4(a)(2) are codified as 29 U.S.C. § 623(a)(1) and (a)(2).

76. *Id.*

77. *Id.* (quoting 29 U.S.C. § 623(a)(2) (2000) (emphasis added by Justice Stevens)).

"incongruity between the employer's actions . . . and the individual employee who adversely suffers because of those actions."[78] In concluding that the text shows that disparate impact claims are allowed, he stated that "an employer who classifies his employees without respect to age may still be liable under the terms of this paragraph if such classification adversely affects the employee because of that employee's age—the very definition of disparate impact."[79]

Justice Stevens then addressed the confusion that *Hazen Paper Co. v. Biggins*[80] caused among the federal appellate courts.[81] Prior to *Hazen Paper*, the lower federal courts had "uniformly interpreted" the ADEA to authorize disparate impact recovery.[82] After *Hazen Paper*, some lower courts "concluded that the ADEA did not authorize a disparate-impact theory of liability."[83] Justice Stevens, addressing those doubts of the availability of disparate impact claims under the ADEA, asserted that nothing in *Hazen Paper* precludes a finding that the ADEA authorizes disparate impact claims.[84]

After finishing his textual analysis, Justice Stevens addressed the RFOA provision's effect on the issue.[85] He asserted that the RFOA provision actually supported the availability of disparate impact claims under the ADEA[86] and used the *Hazen Paper* opinion to demonstrate that the Court had already discussed the RFOA provision's effect on the availability of disparate impact claims.[87] He stated that *Hazen Paper* made it clear that a disparate impact claim cannot exist when the RFOA provision applies.[88]

With its inclusion of the RFOA provision, Congress was attempting to establish a standard for deciding which disparate impact claims would be permissible under the ADEA and which claims would not.[89] Congress could have prevented all disparate impact claims under the ADEA by not including

---

78. *Id.*
79. *Id.*
80. 507 U.S. 604 (1993).
81. *Smith*, 125 S. Ct. at 237 (Stevens, J., plurality opinion).
82. *Id.*
83. *Id.*
84. *Id.* at 238.
85. *Id.*
86. *Id.* at 238-39.
87. *Id.* at 238.
88. *Id.* ("'[T]here is no disparate impact treatment under the ADEA when the factor motivating the employer is some feature other than the employee's age.'" (alteration in original) (quoting *Hazen Paper*, 507 U.S. at 609)). Relying on *Hazen Paper*, Justice Stevens stated, "In most disparate-treatment cases, if an employer in fact acted on a factor other than age, the action would not be prohibited under subsection (a) in the first place." *Id.* Additionally, the RFOA provision serves to preclude liability if the negative disparate impact was based on a factor, other than age, that was "reasonable." *Id.*
89. *See id.*

the RFOA provision, as it did with the Equal Pay Act.[90] Because Congress did include the RFOA, it intended for disparate impact claims to be available under the ADEA.[91]

After addressing the RFOA provision, Justice Stevens discussed Department of Labor and EEOC interpretations of the issue.[92] Because the Department of Labor initially drafted the legislation and the EEOC was responsible for implementing the statute through regulations, their interpretations are significant.[93] He found that both agencies' interpretations implied the availability of a disparate impact claim under the ADEA.[94]

In the last section of his opinion, Justice Stevens stated that "the scope of disparate-impact liability under the ADEA is narrower than under Title VII."[95] He based his opinion on two textual differences, the RFOA provision found only in the ADEA and the Civil Rights Act of 1991 that expanded disparate impact coverage only for Title VII.[96] He attributed the narrower scope of disparate impact under the ADEA to "the fact that age, unlike race or other classifications protected by Title VII, not uncommonly has relevance to an individual's capacity to engage in certain types of employment."[97]

Turning to the facts of the *Smith* case, the Court upheld the lower court's ruling that no valid disparate impact claim was alleged.[98] The Court relied on its decision in *Wards Cove* and stated that, to prevail, the employee had to "'isolat[e] and identify[] *specific* employment practices'" that resulted in disparate impact.[99] In this case, the plaintiffs failed to fulfill this requirement.[100] The Court reasoned that because the City's actions were covered by the RFOA provision, the City was not subject to a disparate impact

---

90. *Id.* at 239 n.11. The Equal Pay Act of 1963 contains a provision that "barred recovery if a pay differential was based 'on any other factor'—reasonable or unreasonable—'other than sex.'" *Id.* (quoting 29 U.S.C. § 206(d)(1)(2000)). Interestingly, the Seventh Circuit reasoned that disparate impact claims were *not* available under the ADEA because of similarities between the RFOA provision and the Equal Pay provision. Holmes, *supra* note 29, at 312.

91. *See Smith*, 544 U.S. 239 n.11 (Stevens, J., plurality opinion).

92. *Id.* at 239

93. *Id.*

94. *Id.* Justice Stevens stated that "[t]he initial regulations . . . permitted [disparate impact] claims if the employer relied on a factor that was not related to age." *Id.*

95. *Id.* at 240(majority opinion).

96. *Id.* Because Congress chose not to amend the ADEA when it amended Title VII, "*Wards Cove*'s pre-1991 interpretation of Title VII's identical language remains applicable to the ADEA." *Id.*

97. *Id.*

98. *Id.* at 241. The Court of Appeals held that there was no disparate impact claim in this case because such claims were "categorically unavailable under the ADEA." *Id.* at 231.

99. *Id.* at 241 (emphasis added) (quoting Wards Cove Packing Co. v. Atonio, 409 U.S. 642, 656 (1989)).

100. *Id.*

claim.[101] The Court stated that "the City's decision to grant a larger raise to lower echelon employees for the purpose of bringing salaries in line with that of surrounding police forces was a decision based on a 'reasonable factor other than age' that responded to the City's legitimate goal of retaining police officers."[102] Although the Court disagreed with the Fifth Circuit's rationale, it nevertheless affirmed the lower court's judgment.[103]

Justice Scalia, writing alone, concurred in part and concurred in the judgment.[104] Applying a slightly different approach from the majority, Justice Scalia called this "an absolutely classic case for deference to agency interpretation."[105] He stated that the EEOC had stated in various regulations and appeared as a party and as an amicus curiae in numerous cases to defend its position that disparate impact claims are cognizable under the ADEA.[106] Justice Scalia spent virtually the rest of his concurrence combating Justice O'Connor's attack of his position.[107] He looked to the text of the EEOC regulation and of the RFOA to rebut her assertions that the EEOC had not taken a position on this issue and that any EEOC position would not deserve deference.[108] In his conclusion, Justice Scalia stated that "[t]he EEOC has express authority to promulgate rules and regulations interpreting the ADEA" and that "[i]t has exercised that authority to recognize disparate-impact claims."[109]

Justice O'Connor, joined by Justices Kennedy and Thomas, concurred only in the judgment.[110] She affirmed the Court of Appeals' decision but disagreed with the majority's holding that permits disparate impact claims under the ADEA.[111] She stated that the ADEA's text, legislative history, and

---

101. *Id.* at 242.

102. *Id.* at 243.

103. *Id.*

104. *Id.* (Scalia, J., concurring in part and concurring in the judgment).

105. *Id.* Justice Scalia based his opinion on section 9 of the ADEA, which "confers upon the EEOC authority to issue 'such rules and regulations as it may consider necessary or appropriate for carrying out the [sic]' ADEA." *Id.* (quoting 29 U.S.C. § 628 (2000)). Justice Scalia applied the *Chevron* standard of deference to agency interpretations. Chevron U.S.A. Inc. v. Natural Res. Def. Council, Inc., 467 U.S. 837, 843 (1984) ("[I]f the statute is silent or ambiguous with respect to the specific issue, the question for the court is whether the agency's answer is based on a permissible construction of the statute.").

106. *Smith*, 544 U.S. at 244 & n.1 (Scalia, J., concurring in part and concurring in the judgment).

107. *See id.* at 244-47.

108. *Id.* at 245. Justice Scalia sought to rebut Justice O'Connor's assertion that "it makes little sense to attribute to the [EEOC] a construction of the relevant statutory text that the agency itself has not actually articulated so that we can then 'defer' to that reading." *Id.* at 266 (O'Connor, J., concurring in the judgment).

109. *Id.* at 247 (Scalia, J., concurring in part and concurring in the judgment).

110. *Id.* (O'Connor, J., concurring in the judgment).

111. *See id.*

purposes made it apparent that disparate impact claims were not authorized under the statute.[112] While Justice Stevens focused on the textual differences between sections 4(a)(1) and 4(a)(2) to conclude that the latter authorized disparate impact claims,[113] Justice O'Connor pointed to the parallels between the two subsections to conclude that the "because of . . . age" language made discriminatory intent a requirement under the ADEA as a whole.[114] She further asserted that the placement of a comma before the "because of . . . age" provision in subsections 4(a)(1) and 4(a)(2) made that language apply to everything before it, regardless of the plurality or singularity of the nouns preceding it.[115] Justice O'Connor concluded that the plurality's textual interpretation was incorrect because the comma preceding the "because of . . . age" provision made it clear that discriminatory intent was required, regardless of sentence structure earlier in the subsection.[116]

Justice O'Connor further noted that the RFOA provision served as "an independent *safe harbor* from liability" for employers who are using facially neutral criteria other than age.[117] She asserted that the RFOA provision "expresses Congress' clear intention that employers *not* be subject to liability absent proof of intentional age-based discrimination."[118] The plurality's requirement that the nonage factor be "reasonable" did not demonstrate Congress's establishment of disparate impact claims under the ADEA.[119] That requirement simply serves to prevent employers from bringing unreasonable justifications for their actions.[120] Because "[r]eliance on an unreasonable nonage factor would indicate that the employer's explanation is, in fact, no more than a pretext for *intentional* discrimination,"[121] the RFOA provision is not evidence that disparate impact claims are available under the ADEA.[122]

Turning to the legislative history, Justice O'Connor examined the Wirtz Report, "[b]ecause the ADEA was modeled on the Wirtz Report's findings and

---

112. *Id.* at 248.

113. *Id.* at 236 & n.6 (Stevens, J., plurality opinion).

114. *Id.* at 248-49 (O'Connor, J., concurring in the judgment) (quoting 29 U.S.C. 623(a)(1), (2) (2000)).

115. *Id.* at 250.

116. *Id.*

117. *Id.* at 252.

118. *Id.* at *251* Justice O'Connor based her assertion on the Court's reasoning in *EEOC v. Wyoming*, 460 U.S. 226 (1983). *Smith*, 544 U.S. at 251 (O'Connor, J., concurring in the judgment). Congress included the RFOA provision "to insure that employers were permitted to use neutral criteria not directly dependant on age, and in recognition of the fact that even criteria that are based on age are occasionally justified." *EEOC v. Wyoming*, 460 U.S. at 232-33.

119. *Smith*, 544 U.S. at 251 (O'Connor, J., concurring in the judgment).

120. *Id.*

121. *Id.* at 253.

122. *See id.*

recommendations."[123] Her opinion noted that the Wirtz Report clearly distinguished age discrimination from the Title VII types of discrimination: "whereas ability is nearly always completely unrelated to the characteristics protected by Title VII, the [Wirtz] Report found that, in some cases," a relationship exists between an individual's age and his job performance.[124] She also noted that the Wirtz Report distinguished between disparate treatment and disparate impact;[125] the Wirtz Report recommended that an age discrimination statute prohibit disparate treatment but address disparate impact claims through "noncoercive measures."[126] The ADEA "draws a clear distinction between 'the setting of arbitrary age limits regardless of potential for job performance' and 'certain otherwise desirable practices [that] may work to the disadvantage of older persons.'"[127] That structure explains why the ADEA differs somewhat from Title VII and why disparate impact claims should be treated differently under each Act.[128] Justice O'Connor also found it significant that the legislative history did not mention disparate impact claims. She noted that the primary focus of anti-discrimination law, at the time the ADEA was passed, was on intentional, not circumstantial, discrimination.[129] She concluded this section by noting that Congress's choice not to allow disparate impact claims under the ADEA was not surprising in light of the historical and present-day differences between Title VII and age discrimination.[130]

Justice O'Connor then disputed the plurality's application of *Griggs* to the ADEA and inference that Congress intended the similar language in the ADEA and Title VII to have identical interpretations.[131] Most notably, *Griggs* was decided after the ADEA was enacted; the case did not put Congress on notice "that language at issue here would be read to authorize disparate impact claims."[132] Similar language can be used to mean different things, even within

---

123. *Id.*

124. *Id.* at 254-55.

125. *Id.* at 255.

126. *Id.* at 256. These "noncoercive measures" include, "programs to increase the availability of employment; continuing education, and adjustment of pension systems, workers' compensation, and other institutional arrangements." *Id.*

127. *Id.* at 257 (alteration in original) (quoting 29 U.S.C. § 621(a)(2) (2000)).

128. *See id.* at 256-58.

129. *Id.* at 258.

130. *See id.* Some of the differences that Justice O'Connor pointed out are given below:

> [(1)] older workers have [not] suffered disadvantages as a result of entrenched historical patterns of discrimination, like racial minorities have, . . . [(2) that] there often *is* a correlation between an individual's age and her ability to perform a job, . . . [and (3) that] many employer decisions that are intended to cut costs or respond to market forces will likely have a disproportionate effect on older workers.

Id. at 258-59.

131. *Id.* at 259 (citing plurality opinion at 233-40).

132. *Id.* at 260.

the same statute;[133] the Court has "not hesitated to give a different reading to the same language . . . if there is strong evidence that Congress did not intend the language to be used uniformly."[134] Justice O'Connor supported her position that Congress only intended to protect against intentional age discrimination by noting the textual and structural differences between the ADEA and Title VII, such as the ADEA's RFOA provision.[135] Justice O'Connor concluded that *Griggs* is inapplicable to the ADEA: "[T]he Court in *Griggs* reasoned that disparate impact liability was necessary to achieve Title VII's ostensible goal of eliminating the cumulative effects of historical racial discrimination . . . , that rationale finds no parallel in the ADEA context."[136]

Justice O'Connor disputed the plurality's and Justice Scalia's assessments of the EEOC and the Department of Labor interpretations of the ADEA.[137] She stated that the 1968 Labor Department bulleting to which the plurality "alludes" does not authorize disparate impact claims.[138] Rather, they only "establish[ ] 'a nonexclusive objective test for employers to use in determining whether they could be certain of qualifying for the' RFOA exemption."[139] She criticized Justice Scalia's application of *Chevron* to the *Smith* case and the "dubious chain of inferences" that led to his conclusion that the EEOC authorized disparate impact claims under the ADEA.[140] Justice O'Connor concluded her dissent by stating that, while she does not believe that disparate impact claims are cognizable under the ADEA, if they are permitted, "they are strictly circumscribed by the RFOA exemption"[141] and are "governed by the standards set forth in . . . *Wards Cove*."[142]

IV. THE CONSEQUENCES OF *SMITH*

Smith holds "that the ADEA . . . authorize(s) 'disparate impact' cases comparable to *Griggs*."[143] The RFOA provision makes the scope of disparate impact more limited under the ADEA; *Wards Cove* governs disparate impact

---

133. *Id.* at 260-61 *see* United States v. Cleveland Indians Baseball Co., 532 U.S. 200, 213 (2001) ("Although we generally presume that 'identical words used in different parts of the same act are intended to have the same meaning,' the presumption 'is not rigid,' and 'the meaning [of the same words] may well vary to meet the purposes of the law, [sic]'" (alteration in original) (quoting Atl. Cleaners & Dryers v. United States, 286 U.S. 427, 433 (1932))).

134. *Smith*, 544 U.S. at 260-61 (O'Connor, J., concurring in the judgment).

135. *Id.* at 261-62.

136. *Id.* at 262.

137. *See id.* at 262-67.

138. *Id.* at 263.

139. *Id.* (quoting Pub. Employees Ret. Sys. of Ohio v. Betts, 492 U.S. 158, 172 (1989)).

140. *Id.* at 264-65.

141. *Id.* at 267.

142. *Id.*

143. *Id.* at 232 (majority opinion).

claims under the ADEA because Congress chose not to amend the ADEA subsequent to that holding.[144] *Smith* clears up any confusion caused by the Court's holding in *Hazen Paper*, a case which resulted in a split among the Courts of Appeals about the availability of ADEA disparate impact claims.[145]

By phrasing the issue as it did,[146] the plurality expanded the scope of this case well beyond the facts particular to *Smith*. Nevertheless, the Court did not provide much guidance about what a successful disparate impact claim under the ADEA must entail. Because the Court affirmed the judgment dismissing the claim,[147] the opinion does not provide a concrete example of a successful ADEA disparate impact claim. It attempts to provide such an example by harkening back to the Wirtz Report, which stated that requiring a high school diploma could have a disparate impact on older workers whose experience gives them the equivalent of a high school diploma.[148]

The Court's guidance that the availability of a disparate impact claim is narrower under the ADEA than under Title VII does not adequately resolve the difficulty of ascertaining the nature of a successful disparate impact claim under the ADEA.[149] For example, switching from traditional pension plans to cash-balance pension plans may have a disparate impact on older employees.[150] The Court's strict limitations on *Smith*'s holding leave a question as to whether such a claim would be successful under the ADEA. The employer could say that it was not making the decision based on age but was solely considering the fiscal consequences; the RFOA provision would

---

144. *Id.* at 240.

145. *See supra* note 53 and accompanying text. The Court recognized the confusion *Hazen Paper* caused but contended that that case "did not address or comment on the issue we decide today." *Smith*, 544 U.S. at 237 (Stevens, J., plurality opinion).

146. The Court phrased the issue in *Smith* as "whether the 'disparate-impact' theory of recovery announced in *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971), for cases brought under Title VII of the Civil Rights Act of 1964, is cognizable under the ADEA." *Smith*, 544 U.S. at 230 (parallel citations omitted).

147. *Smith*, 544 U.S. at 232.

148. *Id.* at 235 n.5 (Stevens, J., plurality opinion) ("'Any formal employment standard which requires, for example, a high school diploma will obviously work against many older workers–unfairly if, despite his limited schooling, an older worker's years of experience have given him the relevant equivalent of a high school education.'" (quoting Wirtz Report, *supra* note 23, at 21)).

149. Soon after the decision in *Smith*, Andrea B. Short wrote: "In allowing a narrower scope of liability under the disparate impact theory under the ADEA than under Title VII, courts have correctly discerned Congress's differing intents in enacting the two nondiscrimination statutes." Andrea B. Short, *Discriminating Among Discrimination: The Appropriateness of Treating Reverse Age Discrimination Differently From Reverse Race Discrimination*, 83 N.C. L. REV. 1065, 1084 (2005).

150. Holmes, *supra* note 29, at 299. See *supra* note 56 and accompanying text for a discussion of cash-balance pension plans.

likely prohibit liability.[151] Additionally, the high degree of proof required under *Wards Cove* may also inhibit the older employees' ability to recover under the ADEA on disparate impact claims.[152]

While *Smith* did help decide an unclear issue, it did not provide us with a clear example of a successful disparate impact claim. The Court, in exercising judicial restraint, kept the scope of liability under disparate impact as narrow as possible. The Courts of Appeals must decide what facts warrant liability under an ADEA disparate impact claim. As a result, the Court will likely be called upon in the future to broaden its holding, or Congress will pass legislation altering this holding, as it did with *Wards Cove* and Title VII.

V. CONCLUSION

In *Smith*, the Court sought to decide whether disparate impact claims are cognizable under the ADEA.[153] Although the Court conceded that such claims were authorized under the ADEA, it limited the scope of such claims. Even Justice O'Connor's dissent, while disagreeing that disparate impact claims are available, stated that if permitted, disparate impact claims must be constrained by the RFOA provision and the Court's holding in *Wards Cove*.[154]

The plurality's failure to give a modern, concrete example of a successful disparate impact claim under the ADEA may cause the Court to have to hear another case regarding this issue in order to shed light on how its limitations are to be applied. Despite its somewhat confusing plurality opinion, *Smith* authorizes lower courts to hear disparate impact claims under the ADEA[155] and makes it clear that *Hazen Paper* does not prevent such claims.[156] The Court's decision in *Smith* allows plaintiffs to continue bringing claims alleging disparate impact based on age and helps older employees exercise their right to equal opportunity for employment. The ability to exercise that right will become increasingly important as medical technology advances and people's life expectancies increase. As a result, the *Smith* decision may facilitate older Americans' transition into longer, more varied careers.

MELISSA M. MARTIN

---

151. *See Smith*, 544 U.S. at 239 (Stevens, J., plurality opinion).

152. *Id.* at 241 ("[T]he employee is 'responsible for isolating and identifying the *specific* employment practices that are allegedly responsible for any observed statistical disparities.'" (emphasis added by Stevens, J.) (quoting Wards Cove Packing Co. v. Atonio, 490 U.S. 642, 656 (1989))). Plaintiffs are under the *Wards Cove* standard unless Congress decides to amend the ADEA, as it did with Title VII.

153. *Smith*, 544 U.S. at 236.

154. *Id.* at 267 (O'Connor, J., concurring in the judgment).

155. *Id.* at 232 (majority opinion).

156. *Id.* at 238 (Stevens, J., plurality opinion).

CIVIL RIGHTS--TITLE IX--AN INDIVIDUAL MAY MAINTAIN A PRIVATE RIGHT OF ACTION UNDER TITLE IX WHEN THE FEDERAL FUNDING RECIPIENT RETALIATES AGAINST THE INDIVIDUAL DUE TO HIS COMPLAINTS ABOUT SEX DISCRIMINATION

JACKSON V. BIRMINGHAM BD. OF EDUC., 544 U.S. 167 (2005).

I. INTRODUCTION

Roderick Jackson, a long-time employee of the Birmingham school district,[1] "brought suit against the Birmingham Board of Education (Board) alleging that the Board retaliated against him because he had complained about sex discrimination in the high school's athletic program."[2] While working as the girls' basketball coach at Ensley High School,[3] Jackson noticed that his team was given neither equal funding nor equal access to athletic facilities and equipment.[4] After Jackson repeatedly complained to his supervisors regarding this inequity,[5] no action was taken; "[i]nstead Jackson began to receive negative work evaluations."[6] Ultimately, he was dismissed from the coaching position.[7]

Shortly after losing his coaching position, Jackson filed suit in the United States District Court for the Northern District of Alabama.[8] He alleged "that the Board violated Title IX by retaliating against him for protesting the discrimination against the girls' basketball team."[9] The Board filed a motion to dismiss, arguing that the complaint failed to state a claim upon which relief could be granted.[10] The Board's motion also contended that plaintiff lacked standing to bring a claim under Title IX "because he was not the victim of the

---

1. The Birmingham Board of Education hired Jackson in 1993 "to serve as a physical education teacher and girls' basketball coach." Jackson v. Birmingham Bd. of Educ., 544 U.S. 167, 171 (2005).

2. *Id.*

3. *Id.* ("Jackson was transferred to Ensley High School in August 1999.").

4. *Id.*

5. *Id.* Jackson began making complaints in December 2000. *Id.*

6. *Id.* at 171-72.

7. *Id.* at 172. Jackson was dismissed from the coaching position in May 2001. *Id.* However, he retained his position as a physical education teacher with Ensley High School. *Id.*

8. *Id.*

9. *Id.*

10. Jackson v. Birmingham Bd. of Educ., No. CV-01-TMP-1866-S, 2002 U.S. Dist. LEXIS 27596, at *1 (N.D. Ala. Jan. 10, 2002).

alleged gender discrimination."[11] The district court granted the Board's motion to dismiss.[12] On appeal, the Eleventh Circuit affirmed the dismissal, holding that a private right of action may not be implied under Title IX when an individual suffers retaliation due to his complaints regarding sex discrimination suffered by others.[13] On certiorari to the United States Supreme Court, *held*, reversed and remanded.[14] Under Title IX, an individual may maintain a private right of action when "the funding recipient retaliates against an individual because he has complained about sex discrimination."[15] *Jackson v. Birmingham Bd. of Educ.*, 544 U.S. 167 (2005).

II. THE DEVELOPMENT OF TITLE IX'S PRIVATE RIGHT OF ACTION FOR INTENTIONAL SEX DISCRIMINATION

The Supreme Court has recognized that an implied private right of action exists for individuals alleging intentional sex discrimination under section 901 of Title IX.[16] The federal circuit courts of appeals have reached conflicting decisions regarding whether this private right of action encompasses retaliation claims or whether, in the alternative, an individual has standing to enforce the

---

11. *Id.* at *3. The motion also contended that plaintiff's claims were preempted by Title VII of the Civil Rights Act. *Id.* at *1.

12. *Id.* at *9. The court relied on *Holt v. Lewis*, 955 F. Supp. 1385, 1389 (N.D. Ala. 1995), *aff'd*, 109 F.3d 771 (11th Cir. 1997) (unpublished table decision) (TABLE NO. 96-6046), *cert. denied*, 522 U.S. 817 (1997), in which the district court declined to recognize Title IX retaliation claims. *Jackson*, 2002 U.S. Dist. LEXIS 27596, at *8. Although *Holt*, an unpublished decision of the Eleventh Circuit Court of Appeals, is not binding precedent, the *Jackson* court gave *Holt* great weight. *Id.* ("Simply stated . . . the Eleventh Circuit's affirmance of Judge Acker's decision in *Holt v. Lewis*, leaves this court with little option but to reject the argument that Title IX implicitly creates a private cause of action for retaliation." (citation omitted)).

13. Jackson v. Birmingham Bd. of Educ., 309 F.3d 1333, 1347 (11th Cir. 2002). The court relied on *Alexander v. Sandoval*, 532 U.S. 275, 293 (2001), in which the United States Supreme Court declined to imply a private right of action under Title VI of the Civil Rights Act of 1964 to enforce disparate impact regulations promulgated by a federal agency under Title VI's enforcement scheme. *Jackson*, 309 F.3d at 1338.

14. Jackson v. Birmingham Bd. of Educ., 544 U.S.167, 184 (2005).

15. *Id.* at 171.

16. *Id.* at 173 ("More than 25 years ago, . . . we held that Title IX implies a private right of action to enforce its prohibition on intentional sex discrimination." (citation omitted)); Cannon v. Univ. of Chi., 441 U.S. 677, 709 (1979) ("Not only the words and history of Title IX, but also its subject matter and underlying purposes, counsel implication of a cause of action in favor of private victims of discrimination."). In *Sandoval*, the Supreme Court held that individuals do not have standing to enforce disparate impact regulations because the "regulations do not simply apply § 601['s rohibition of intentional discrimination]." *Sandoval*, 532 U.S. at 285. Therefore, individuals may only maintain a private right of action for retaliation if retaliation is a form of intentional sex discrimination prohibited by section 901 of Title IX. *See Jackson*, 544 U.S. at 177-78.

anti-retaliation regulation promulgated by the Department of Education pursuant to section 902 of Title IX.[17] The United States Supreme Court resolved this issue in *Jackson v. Birmingham Board of Education*.[18]

Title IX of the Education Amendments of 1972 was enacted "to avoid the use of federal resources to support discriminatory practices" and "to provide individual citizens effective protection against those practices."[19] Section 901 of Title IX provides that "[n]o 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."[20] Section 902 of Title IX provides that "[e]ach Federal department and agency which is empowered to extend Federal financial assistance to any education program . . . is authorized and directed to effectuate the provisions of section [901] . . . by issuing rules, regulations, or orders."[21]

In construing Title IX, courts have not only considered the statutory language itself but also the legislative history. Because Title IX was modeled after Title VI[22] of the Civil Rights Act of 1964,[23] the Supreme Court has adopted a similar approach in interpreting the two statutes.[24] Therefore, cases construing Title VI are often helpful in predicting how the Court will resolve the same issues arising under Title IX.

Neither Title IX nor Title VI expressly prohibits the funding recipient

---

17. *Jackson*, 544 U.S. at 172-73. *Compare* Jackson v. Birmingham Bd. of Educ., 309 F.3d 1333, 1338 (11th Cir. 2002) ("[W]e hold that neither Title IX itself nor 34 C.F.R. § 100.7(e) [promulgated under § 902] implies a private right of action for retaliation in Jackson's favor."), *with* Lowery v. Texas A & M Univ. Sys., 117 F.3d 242, 252 (5th Cir. 1997) ("We . . . conclud[e] that title IX [sic] affords an implied cause of action for retaliation under 34 C.F.R. § 100.7(e) and that the employees of federally funded educational institutions are members of the class for whose special benefit this provision was enacted."), and Preston v. Virginia *ex rel.* New River Cmty. Coll., 31 F.3d 203, 206 & n.2 (4th Cir. 1994) ("Retaliation against an employee for filing a claim of gender discrimination is prohibited under Title IX.").

18. *Jackson*, 544 U.S. at 171.

19. *Cannon*, 441 U.S. at 704.

20. Title IX of the Education Amendments of 1972 § 901, 20 U.S.C. § 1681(a) (2000).

21. Title IX of the Education Amendments of 1972 § 902, 20 U.S.C. § 1682 (2000).

22. Civil Rights Act of 1964, Title VI § 601, 42 U.S.C. § 2000d (2000) ("No person in the United States shall, on the ground of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance.").

23. *Cannon*, 441 U.S. at 694 n.16; Gebser v. Lago Vista Indep. Sch. Dist., 524 U.S. 274, 286 (1998) (stating that Title VI "is parallel to Title IX except that it prohibits race discrimination, not sex discrimination, and applies in all programs receiving federal funds, not only in education programs.").

24. *Cannon*, 441 U.S. at 696. In fact, "[t]he drafters of Title IX explicitly assumed that it would be interpreted and applied as Title VI had been during the preceding eight years." *Id.* at 696 & n.19.

from retaliating against an individual who complains about discrimination.[25] However, federal agencies have issued regulations prohibiting retaliation pursuant to section 902 of Title IX and section 602, the equivalent section of Title VI.[26] In each of these sections, "Congress created and authorized an elaborate administrative enforcement scheme for Title IX"[27] and Title VI, respectively. As part of this scheme, agencies that disperse federal funds are authorized to promulgate regulations to enforce Titles IX and VI; the agencies may even refuse to grant or continue funding if a recipient fails to comply with the regulations.[28] While individuals may file administrative complaints alleging that a funding recipient has violated a regulation, a federal agency receiving a complaint has complete discretion over any investigation.[29] The individual may not participate in the investigation, and will not receive compensation if the agency uncovers a violation.[30] Hence, the ability to file a private cause of action in federal court is very appealing to victims of discrimination.[31]

In *Cannon v. University of Chicago*,[32] the Supreme Court recognized an implied private right of action under section 901 of Title IX for victims of sex discrimination, even though the statute did not expressly provide for such an action.[33] The Supreme Court considered four factors in determining whether to imply a private right of action under Title IX:

---

25. Both section 601 of Title VI and section 901 of Title IX prohibit acts of intentional discrimination without enumerating specific behaviors tantamount to intentional discrimination. *See* 20 U.S.C. § 1681(a) (2000); 42 U.S.C. § 2000d (2000). *But see* 42 U.S.C. § 2000e-3(a) (2000) (Title VII expressly prohibits retaliation).

26. Civil Rights Act of 1964, Title VI § 602, 42 U.S.C. § 2000d-1 (2000) ("Each Federal department and agency which is empowered to extend Federal financial assistance to any program . . . is authorized and directed to effectuate the provisions of section [601] . . . by issuing rules, regulations, or orders . . . . "). Pursuant to this authority, the Department of Education promulgated an anti-retaliation regulation that states in pertinent part that "[n]o recipient or other person shall intimidate, threaten, coerce, or discriminate against any individual for the purpose of interfering with any right or privilege secured by section 601 of the Act . . . because he has made a complaint." 34 C.F.R. § 100.7(e) (2004). This anti-retaliation regulation, along with the other procedural provisions applicable to Title VI, were incorporated into Title IX by reference. 34 C.F.R. § 106.71 (2004).

27. Jackson v. Birmingham Bd. of Educ., 390 F.3d 1333, 1336 (11th Cir. 2002).

28. 20 U.S.C. § 1682 (2000); 42 U.S.C. § 2000d-1 (2000). In reality, however, "federal funding agencies almost never impose the draconian remedy of terminating funding. Instead, agencies usually require a recipient found guilty of discrimination to sign a binding settlement in which it agrees to end any discriminatory practices." Bradford C. Mank, *Are Anti-Retaliation Regulations in Title VI or Title IX Enforceable in a Private Right of Action: Does* Sandoval *or* Sullivan *Control This Question?*, 35 SETON HALL L. REV. 47, 58 (2004).

29. Mank, *supra* note 28, at 57.

30. *Id.*

31. *Id.*

32. 441 U.S. 677 (1979).

33. *Id.* at 717.

> First, is the plaintiff "one of the class for whose *especial* benefit the statute was enacted" . . . ? Second, is there any indication of legislative intent, explicit or implicit, either to create such a remedy or to deny one? Third, is it consistent with the underlying purposes of the legislative scheme to imply such a remedy for the plaintiff? And finally, is the cause of action one traditionally relegated to state law, in an area basically the concern of the States, so that it would be inappropriate to infer a cause of action based solely on federal law?[34]

In considering the legislative intent of Title IX, the *Cannon* Court observed that "[t]he drafters of Title IX explicitly assumed that it would be interpreted and applied as Title VI had been during the preceding eight years."[35] Therefore, because Title VI had already been construed as creating a private right of action,[36] the Supreme Court presumed "both that those representatives were aware of the prior interpretation of Title VI and that that interpretation reflects their intent with respect to Title IX."[37] With this in mind, the Supreme Court held that an implied private right of action exists under Title IX for victims of sex discrimination.[38]

In subsequent cases, the Supreme Court has "defined the contours of that right of action."[39] In *Guardians Association v. Civil Service Commission*,[40] a deeply divided Supreme Court held that section 601 of Title VI only prohibits intentional discrimination;[41] it does not prohibit disparate impact discrimination.[42] Therefore, to recover under Title VI, an individual must prove that the agency was "motivated by an invidious intent."[43] However, five justices either "stated or strongly implied" that section 602 allows an agency

---

34. *Id.* at 688 & n.9 (internal citations omitted) (quoting Cort v. Ash, 422 U.S. 66, 78 (1975)).

35. *Id.* at 696.

36. *See* Bossier Parish Sch. Bd. v. Lemon, 370 F.2d 847, 852 (5th Cir. 1967). "Most particularly, in 1967, a distinguished panel of the Court of Appeals for the Fifth Circuit squarely decided this issue in an opinion that was repeatedly cited with approval and never questioned during the ensuing five years." *Cannon*, 441 U.S. at 696.

37. *Cannon*, 441 U.S. at 697-98.

38. *Id.* at 709. Since this holding, the Supreme Court has become reluctant to allow a private right of action when it is not expressly provided for in the statute. Mank, *supra* note 28, at 68. Plaintiffs are now required to prove "that Congress intended to authorize remedies for private litigants." *Id.* However, "[t]he Supreme Court has never overruled *Cannon*'s holding." *Id.*

39. Jackson v. Birmingham Bd. of Educ., 544 U.S. 167, 173 (2005).

40. 463 U.S. 582 (1983).

41. *Id.* at 584. *See also* Alexander v. Choate, 469 U.S. 287, 293 (1985) ("[T]he [*Guardians*] Court held that Title VI itself directly reached only instances of intentional discrimination.").

42. *Guardians*, 463 U.S. at 593.

43. *Id.* at 645 (Stevens, J., dissenting).

to enact regulations prohibiting disparate impact discrimination.[44] This implication was confirmed in *Alexander v. Choate*,[45] when the Supreme Court relied on the *Guardians* Court's holding that section 602 authorizes agencies to issue disparate impact regulations.[46]

Although an implied private right of action clearly exists for victims of intentional discrimination under section 601 of Title VI, in *Alexander v. Sandoval*,[47] the Supreme Court held that no private right of action exists to enforce disparate impact regulations promulgated under section 602.[48] Prior to *Sandoval*, courts that implied a private right of action for retaliation under Title IX often did so because of the anti-retaliation regulation promulgated by the Department of Education pursuant to section 902.[49] After *Sandoval*'s holding, courts were required to determine whether this anti-retaliation regulation implicated intentional discrimination or "conduct other than intentional discrimination,"[50] such as disparate impact discrimination. Only if retaliation constituted intentional sex discrimination could a plaintiff, or victim, maintain a private right of action.

Generally, the Supreme Court has adopted a liberal approach when determining what constitutes intentional sex discrimination in violation of section 901. For example, the Court has held that Title IX's private right of action for intentional sex discrimination encompasses a federal funding "recipient's deliberate indifference to a teacher's sexual harassment of a student"[51] and, in some instances, a federal funding recipient's intentional

---

44. Mank, *supra* note 28, at 71.

> Justice Stevens, in his dissenting opinion . . . . [stated], "although the petitioners had to prove that the respondents' actions were motivated by an invidious intent in order to prove a violation of [Title VI], they only had to show that the respondents' actions were producing discriminatory effects in order to prove a violation of [the regulations]."

Id. at 71-72 (quoting *Guardians*, 463 U.S. at 645 (Stevens, J., dissenting)). Justice Stevens was joined by Justices Brennan and Blackmun in this opinion. *Id.* at 71. In addition, "Justices White and Marshall each would have allowed disparate impact suits under either § 601 or § 602." *Id.* at 72.

45. 469 U.S. 287 (1985).

46. *Id.* at 293 ("[T]he [*Guardians*] Court held that actions having an unjustifiable disparate impact on minorities could be redressed through agency regulations designed to implement the purposes of Title VI.").

47. 532 U.S. 275 (2001).

48. *Id.* at 293.

49. 34 C.F.R. § 100.7(e) (2004); 34 C.F.R. § 106.71 (2004); *see e.g.*, Lowrey v. Texas A & M Univ. Sys., 117 F.3d 242, 249 n.9 (5th Cir. 1997) ("[C]oncluding that title IX [sic] affords an implied cause of action for retaliation under 34 C.F.R.§ 100.7(e)."); Preston v. Virginia *ex rel.* New River Cmty. Coll., 31 F.3d 203, 206 & n.2 (4th Cir. 1994) ("Retaliation against an employee for filing a claim of gender discrimination is prohibited under Title IX [34 C.F.R. § 100.7(e)]").

50. Peters v. Jenney, 327 F.3d 307, 316 (4th Cir. 2003).

51. Jackson v. Birmingham Bd. of Educ., 544 U.S. 167, 173 (2005); Gebser v. Lago Vista Indep. Sch. Dist., 524 U.S. 247, 292-93 (1998).

indifference "to sexual harassment of a student by another student."[52] However, for some time it remained unclear how to classify a federal funding recipient's retaliation due to complaints of sex discrimination. Is the retaliation a form of intentional sex discrimination for which a private cause of action is available under section 901? Alternatively, is the retaliation conduct other than intentional discrimination whose only means of enforcement is the administrative scheme of section 902?

Following the Supreme Court's holding in *Sullivan v. Little Hunting Park, Inc.*,[53] courts consistently considered retaliation in other contexts to be a form of intentional discrimination.[54] In *Sullivan*, a white homeowner suffered retaliation for protesting the discrimination against his black tenant.[55] The Court found that the owner had standing to bring a retaliation claim under 42 U.S.C. § 1982,[56] even though the statute did not expressly mention retaliation, and even though he was not the direct victim of the underlying racial discrimination.[57]

However, the Eleventh Circuit Court of Appeals, in *Jackson v. Birmingham Board of Education*,[58] held that Title IX does not imply "a private

---

52. *Jackson*, 544 U.S. at 173; Davis v. Monroe County Bd. of Educ., 526 U.S. 629, 643 (1999).

53. 396 U.S. 229 (1969).

54. Mank, *supra* note 28, at 82; *see also* Foley v. Univ. of Houston Sys., 324 F.3d 310, 315-16 (5th Cir. 2003) ("[A]n employee's claim that he was subjected to retaliation because he complained of race discrimination is a cognizable claim under § 1981(b)."); Johnson v. Univ. of Cincinnati, 215 F.3d 561, 575 (6th Cir. 2000) (holding that an employee who speaks out on behalf of minorities may have a viable claim under 42 U.S.C. § 1981 for retaliation notwithstanding employee's own race or ethnicity); Andrews v. Lakeshore Rehab. Hosp., 140 F.3d 1405, 1411-13 (11th Cir. 1998) (holding that, after the 1991 revision to § 1981, retaliation claims are still cognizable); Phelps v. Wichita Eagle-Beacon, 886 F.2d 1262, 1267 (10th Cir. 1989) ("[A]lleged discrimination against a white person because of his association with blacks may state a cause of action under Section 1981."); Skinner v. Total Petroleum, Inc., 859 F.2d 1439, 1447 (10th Cir. 1988) ("[A]n employee who has been the subject of employer retaliation because of his efforts to vindicate the rights of racial minorities may bring an action under § 1981."); Fiedler v. Marumsco Christian Sch., 631 F.2d 1144, 1150 (4th Cir. 1980) (holding that § 1981 affords a remedy for the retaliatory expulsion of a student from a private sectarian school); Chandamuri v. Georgetown Univ., 274 F. Supp. 2d 71, 83 (D.D.C. 2003) ("[A]n action against retaliation is implicitly within the scope of Title VI's prohibition on intentional discrimination.").

55. *Sullivan*, 396 U.S. at 234-35. Sullivan was expelled from a "corporation organized to operate a community park and playground facilities" when he protested the board's refusal to recognize the assignment of membership rights to his tenant. *Id.*

56. 42 U.S.C. § 1982 (1969) ("All citizens of the United States shall have the same right, in every State and Territory, as is enjoyed by white citizens thereof to inherit, purchase, lease, sell, hold, and convey real and personal property.").

57. *Sullivan*, 396 U.S. at 237 ("[T]here can be no question but that Sullivan has standing to maintain this action.").

58. 309 F.3d 1333 (11th Cir. 2002).

right of action in favor of individuals who, although not themselves the victims of gender discrimination, suffer retaliation because they have complained about gender discrimination suffered by others."[59] Like the Supreme Court in *Sandoval*, the Eleventh Circuit focused on the congressional intent in enacting Title IX to determine whether a private right of action should be implied.[60]

In its analysis, the circuit court first examined the text of section 901.[61] Finding no mention of retaliation in the statute itself,[62] the court concluded that this "weighs powerfully against a finding that Congress intended Title IX to reach retaliatory conduct."[63] The court noted that "the fact that Congress felt required to prohibit retaliation expressly under Title VII [of the Civil Rights Act of 1964[64]] may indicate that Congress did not intend the concept of 'discrimination' in Title IX to be read sufficiently broadly to cover retaliation."[65]

The Eleventh Circuit then turned to section 902 of Title IX.[66] Relying on *Sandoval*'s holding that no private right of action exists to enforce disparate impact regulations,[67] the court concluded that "[s]ection 902 plainly does not disclose any congressional intent to imply a private right of action of any kind, let alone against retaliation."[68] Therefore, the Eleventh Circuit concluded that the Department of Education's anti-retaliation regulation[69] provided no private remedy to the appellant.[70] As an alternative to the reasoning of *Sandoval*, the Eleventh Circuit stated that "even if Title IX did aim to prevent and remedy retaliation for complaining about gender discrimination, [appellant] is plainly

---

59. *Id.* at 1335.

60. *Id.* at 1344-47.

61. *Id.* at 1344-45.

62. *Id.* at 1344 ("Nothing in the text indicates any congressional concern with retaliation that might be visited on those who complain of Title IX violations. Indeed, the statute makes no mention of retaliation at all.").

63. *Id.* at 1345.

64. Civil Rights Act of 1964, Title VII § 704, 42 U.S.C. § 2000e-3(a) (2000) ("It shall be an unlawful employment practice for an employer to discriminate against any of his employees . . . because he has opposed any practice made an unlawful employment practice by this subchapter.").

65. *Jackson*, 309 F.3d at 1344 n.12.

66. *Id.* at 1345.

67. *See* Alexander v. Sandoval, 532 U.S. 275, 284 (2001).

> We do not doubt that regulations applying § 601's ban on intentional discrimination are covered by the cause of action to enforce that section. . . . A Congress that intends the statute to be enforced through a private cause of action intends the authoritative interpretation of the statute to be so enforced as well.

Id. However, regulations that go beyond what the statute itself prohibits are not enforceable through a private right of action. *Id.*

68. *Jackson*, 309 F.3d at 1345.

69. *See supra* note 26 and accompanying text.

70. *Jackson*, 309 F.3d at 1346.

not within the class meant to be protected by Title IX."[71]

Conversely, in *Peters v. Jenney*,[72] the Fourth Circuit Court of Appeals held that a private action for retaliation is available under Title VI.[73] In its attempt to determine whether *Sandoval* precluded such an action, the court noted that it

> must resolve the question of whether 34 C.F.R. § 100.7(e)'s retaliation prohibition [promulgated by the Department of Education] is an interpretation of § 601's core antidiscrimination mandate. If § 100.7(e) is an interpretation of § 601 that is valid under *Chevron*,[74] it commands deference and may be enforced via an implied private right of action. If instead, § 100.7(e) is a regulation which, even if valid as a § 602 "means of effectuating" Title VI, nonetheless "forbid[s] conduct that § 601 permits," namely conduct other than intentional discrimination, the regulation may not be enforced via an implied private right of action.[75]

The court, relying on *Sullivan* and its progeny, asserted that intentional discrimination encompasses "retaliation against those who oppose the prohibited discrimination," even if they were not direct victims of the discrimination.[76] Therefore, regulations prohibiting retaliation may give rise to a private right of action if the complainant alleges that the retaliation resulted from complaints about intentional discrimination prohibited by section 601.[77] If, on the other hand, the complainant alleges that the retaliation resulted from complaints regarding disparate impact discrimination, no private action is available under *Peters*.[78]

III. JACKSON V. BIRMINGHAM BOARD OF EDUCATION: TITLE IX'S PRIVATE RIGHT OF ACTION ENCOMPASSES RETALIATION CLAIMS

In *Jackson v. Birmingham Board of Education*,[79] a five-four decision, the Supreme Court held that Title IX's private right of action for intentional sex

---

71. *Id.* As Mank notes, "*Jackson*'s limitation of retaliation actions to actual victims of discrimination is contrary to *Sullivan* and its progeny." Mank, *supra* note 28, at 94.

72. 327 F.3d 307 (4th Cir. 2003).

73. *Id.* at 310.

74. *Id.* "Under the . . . *Chevron* standard, 'when it appears that Congress delegated authority to an agency generally to make rules carrying the force of law, we give great deference to an administrative implementation of the particular statutory provision.'" *Peters*, 327 F.3d at 315-16 (internal quotation marks and alterations omitted) (quoting McDaniels v. United States, 300 F.3d 407, 411 (4th Cir. 2002)).

75. *Peters*, 327 F.3d at 316 (alteration in original) (citation omitted) (quoting Alexander v. Sandoval, 532 U.S. 275, 285 (2001)).

76. *Id.* at 317-18.

77. *Id.* at 319.

78. *Id.*

79. 544 U.S. 167 (2005).

discrimination encompasses retaliation claims.[80] Writing for the majority, Justice O'Connor delivered the opinion of the Court.[81] After summarizing the various forms of intentional sex discrimination encompassed by Title IX's private right of action,[82] the Court determined that retaliation against a person who has complained of sex discrimination also falls within this implied right of action.[83] The Court reasoned that "[r]etaliation is, by definition, an intentional act. It is a form of 'discrimination' because the complainant is being subjected to differential treatment . . . . Moreover, retaliation is discrimination 'on the basis of sex' because it is an intentional response to the nature of the complaint: an allegation of sex discrimination."[84] Therefore, retaliation constitutes intentional sex discrimination prohibited by section 901 of Title IX.[85]

The Supreme Court admonished the Eleventh Circuit for its conclusion that the failure to mention retaliation expressly in the statute necessarily means that section 901 of Title IX does not prohibit such conduct.[86] The Court used sexual harassment as an example of a form of intentional sex discrimination not expressly mentioned in the statute but held to be within its scope.[87] The Court then rejected the Board of Education's related argument that because Congress prohibited retaliation expressly in Title VII, its failure to do so in Title IX indicates that it did not intend to prohibit retaliation in this statute.[88] In comparing Title VII with Title IX, the Court pointed out that "Title VII . . . is a vastly different statute from Title IX."[89] Title VII's cause of action is express, while Title IX's is implied; further, Title VII specifically lists the conduct that constitutes discrimination, while Title IX is a broad prohibition on discrimination.[90] "Because Congress did not list *any* specific discriminatory practices when it wrote Title IX, its failure to mention one such practice does not tell us anything about whether it intended that practice to be covered."[91]

The Court then pointed out that Title IX was enacted merely three years after *Sullivan* was decided; therefore, the Court determined that it was safe to assume that Congress was aware of *Sullivan* and expected Title IX to be interpreted in conformity with *Sullivan*'s holding.[92] The Court reasoned that

80. *Id.* at 169, 171.
81. *Id.* at 171.
82. *Id.* at 173. *See supra* notes 51-52 and accompanying text.
83. *Jackson*, 544 U.S. at 173.
84. *Id.* at 173-74 (citations omitted).
85. *Id.* at 174.
86. *Id.*
87. *Id.*
88. *Id.* at 175.
89. *Id.*
90. *Id.*
91. *Id.*
92. *Id.* at 176.

"[r]etaliation for Jackson's advocacy of the rights of the girls' basketball team in this case is 'discrimination' 'on the basis of sex,' just as retaliation for advocacy on behalf of a black lessee in *Sullivan* was discrimination on the basis of race."[93]

Next, the Court addressed the Board's contention that its decision in *Sandoval* requires a holding that no private cause of action exists for retaliation.[94] The Court determined that *Sandoval*'s holding, that no private right of action exists to enforce disparate impact regulations promulgated under section 602, was inapplicable to the case at hand.[95] The Court's holding does *not* rely on the anti-retaliation regulation promulgated by the Department of Education under section 902.[96] Instead, the Court held that the text of Title IX itself implicitly prohibits retaliation in its blanket prohibition of intentional discrimination on the basis of sex.[97] Since retaliation is a form of intentional sex discrimination, there was no need to resort to the enforcement mechanisms of section 902.[98] Thus, a private action for retaliation may be maintained under the provisions of section 901.[99]

The Court referred to section 901 and *Sullivan* to refute the Board's contention that petitioner is not entitled to maintain a private right of action for retaliation because he is an "indirect victi[m]" of sex discrimination.[100] The Court noted that the statute "does not require that the victim of the retaliation must also be the victim of the discrimination that is the subject of the original complaint . . . . Where the retaliation occurs because the complainant speaks out about sex discrimination, the 'on the basis of sex' requirement is satisfied."[101] Similarly, "*Sullivan* made clear that retaliation claims extend to those who oppose discrimination against others."[102]

In addition, the Court noted the importance of allowing individuals to maintain private actions for retaliation.[103] One motivation for enacting Title IX was to protect individuals against sex discrimination.[104] If federal funding recipients were able to retaliate freely against individuals who complain of sex discrimination, those individuals would be less likely to report discriminatory conduct.[105] As a result, "Title IX's enforcement scheme would unravel."[106]

---

93. *Id.* at 176-77.
94. *Id.* at 177.
95. *Id.* at 178.
96. *Id.*
97. *Id.*
98. *Id.*
99. *Id.*
100. *Id.* at 179 (alteration in original) (quoting Brief for Respondent at 33).
101. *Id.*
102. *Id.* at 180.
103. *Id.*
104. *Id.*
105. *Id.*
106. *Id.*

Reporting sex discrimination is crucial to the enforcement of Title IX; funding recipients must receive "actual notice" of discrimination before an individual may bring a suit under the statute.[107] "If recipients were able to avoid such notice by retaliating against all those who dare complain, the statute's enforcement scheme would be subverted."[108] In addition, "sometimes adult employees are 'the only effective adversar[ies]' of discrimination in schools."[109]

When Congress uses its Spending Clause power[110] to enact a statute, as it did when enacting Title IX, "'private damages actions are available only where recipients of federal funding had adequate notice that they could be liable for the conduct at issue.'"[111] The Board argued that a private action is not available in this case because it lacked notice of potential liability for retaliation claims under Title IX.[112] The Court disagreed, stating that its 1979 holding in *Cannon v. University of Chicago*[113] put the Board on notice that Title IX allowed a private cause of action for intentional sex discrimination.[114] Also, the subsequent holdings in *Gebser v. Lago Vista Independent School District*[115] and *Davis ex rel. LaShonda D. v. Monroe County Board of Education*[116] should have notified the Board that Title IX's private right of action broadly encompasses many forms of intentional sex discrimination.[117] The Court also noted that "the Courts of Appeals that had considered the question at the time of the conduct at issue in this case all had already interpreted Title IX to cover retaliation."[118] Given this context, the Board had

---

107. *Id.* at 180-81.

108. *Id.* at 181.

109. *Id.* (alteration in original) (quoting Sullivan v. Little Hunting Park Inc., 396 U.S. 229, 237 (1969)).

110. "When Congress enacts legislation under its spending power, that legislation is 'in the nature of a contract: in return for federal funds, the States agree to comply with federally imposed conditions.'" *Id.* at 181-82 (quoting Pennhurst State Sch. & Hosp. v. Halderman, 451 U.S. 1, 17 (1981)).

111. *Jackson,* 544 U.S. at 181 (quoting Davis v. Monroe County Bd. of Educ., 526 U.S. 629, 640 (1999)).

112. *Id.* at 182.

113. 113 U.S. 677 (1979).

114. *Jackson,* 544 U.S. at 182.

115. 524 U.S. 274 (1998).

116. 526 U.S. 629 (1999).

117. *Jackson,* 544 U.S. at 183.

118. *Id.*; *see, e.g.*, Lowrey v. Texas A & M Univ. Sys., 117 F.3d 242, 252 (5th Cir. 1997) ("We . . . conclud[e] that title IX [sic] affords an implied cause of action for retaliation . . . and that employees of federally funded educational institutions are members of the class for whose special benefit this provision was enacted."); Preston v. Virginia *ex rel.* New River Cmty. Coll., 31 F.3d 203, 206 (4th Cir. 1994) ("Retaliation against an employee for filing a claim of gender discrimination is prohibited under Title IX.").

notice and should have known that it could be held liable for retaliation.[119] Therefore, the Court concluded that petitioner may maintain a private cause of action for retaliation under Title IX. In order to recover for retaliation, however, petitioner must prove that he was retaliated against "*because* he complained of sex discrimination."[120] The Court remanded the case to determine whether retaliation because of complaints of sex discrimination had occurred.[121]

In his dissenting opinion, Justice Thomas disagreed with the majority's holding that Title IX's general prohibition of intentional sex discrimination encompasses retaliation claims.[122] He also contended that petitioner does not meet Title IX's "on the basis of sex" requirement.[123] Considering the plain meaning of the text, Justice Thomas found that the "on the basis of sex" requirement in Title IX means "on the basis of the plaintiff's sex, not the sex of some other person."[124] Therefore, because petitioner's complaint focused on discrimination suffered by others, Justice Thomas found that petitioner does not belong to the class of people protected by the statute.[125]

In addition, Justice Thomas noted a unique characteristic of retaliation claims; a retaliation claim may succeed even though sex discrimination never actually took place.[126] A person alleging retaliation is required to prove only that he or she complained about sex discrimination and that this complaint resulted in retaliatory conduct.[127] The claimant need not prove the existence of the underlying sex discrimination to recover for retaliation.[128] Therefore, Justice Thomas concluded that "[b]ecause Jackson's claim for retaliation is not a claim that his sex played a role in his adverse treatment, the statute's plain terms do not encompass it."[129]

Justice Thomas also found Congress's failure to explicitly mention retaliation in the statute significant.[130] Noting that Congress explicitly prohibited retaliation in Title VII and other statutes, Justice Thomas concluded that "Congress' [sic] failure to include similar text in Title IX shows that it did

---

119. *Jackson*, 544 U.S. at 183-84.

120. *Id.* at 184.

121. *Id.*

122. *Id.* at 185 (Thomas, J., dissenting) ("A claim of retaliation is not a claim of discrimination on the basis of sex.").

123. *Id.*

124. *Id.*

125. *Id.* at 186-87.

126. *Id.* at 187-88 ("Retaliation . . . cannot be said to be discrimination on the basis of anyone's sex, because a retaliation claim may succeed where no sex discrimination ever took place.").

127. *Id.* at 187.

128. *Id.*

129. *Id.* at 186-87.

130. *Id.* at 189.

not authorize private retaliation actions."[131]

Despite the majority's contrary conclusion, Justice Thomas found that the Board had no notice of its potential liability for retaliation claims under Title IX.[132] He contended that the majority's holding would be detrimental to the notice requirement: "[T]he Court's rationale untethers notice from the statute. The Board, and other Title IX recipients, must now assume that if conduct can be linked to sex discrimination – no matter how attenuated that link – this Court will impose liability under Title IX."[133]

Justice Thomas also found that the majority's opinion conflicted with the Court's standard "for implying causes of action to enforce federal statutes."[134] To determine if a statute implies a private right of action, the Court must consider both the nature of the conduct and the plaintiff.[135] Justice Thomas found no private right of action in this case; Congress provided no explicit prohibition of retaliation in Title IX, and "extending the implied cause of action under Title IX to claims of retaliation expands the class of people the statute protects beyond the specified beneficiaries."[136]

Further, Justice Thomas noted that "[t]he majority's reliance on *Sullivan* . . . is wholly misplaced."[137] He distinguished *Sullivan* by noting that section 1982 (the relevant anti-discrimination statute in *Sullivan*) was enacted pursuant to Congress's Thirteenth Amendment enforcement power rather than its Spending Clause power.[138] He emphasized that in order for the plaintiff in *Sullivan* "[t]o make out his third-party claim on behalf of the black lessee, the white lessor would necessarily be required to demonstrate that the defendant had discriminated against the black lessee on the basis of race."[139] Conversely, under the majority's holding, petitioner in the instant case is not required to prove the existence of sex discrimination to recover for retaliation.[140] Accordingly, Justice Thomas commented that "by recognizing Jackson's claim, the majority creates an entirely new cause of action for a secondary rights holder, beyond the claim of the original rights holder, and well beyond *Sullivan*."[141] Therefore, Justice Thomas "would hold that § 901 [of Title IX] does not encompass private actions for retaliation."[142]

---

131. *Id.*
132. *Id.* at 191-92.
133. *Id.* at 192.
134. *Id.* ("Whether a statute supplies a cause of action is a matter of statutory interpretation.").
135. *Id.*
136. *Id.* at 194.
137. *Id.*
138. *Id.* at 194-95.
139. *Id.* at 194.
140. *Id.*
141. *Id.* at 194-95.
142. *Id.* at 196.

IV. THE IMPACT OF THE *JACKSON* DECISION

The Supreme Court's decision in *Jackson v. Birmingham Board of Education* "is consistent with Congress's central purpose of barring intentional discrimination"[143] and in accord with the Court's prior holdings. Since its decision in *Sullivan*, the Court has broadly defined "intentional discrimination" in Title IX and other contexts.[144] In light of these decisions, it is not surprising that the Court interpreted Title IX's prohibition of intentional sex discrimination to include retaliation for complaints of sex discrimination.

While this broad definition of intentional discrimination is consistent with prior holdings, it is somewhat difficult to reconcile with the "actual notice" requirement of Spending Clause legislation. As Justice Thomas points out, because funding recipients must have actual notice of their potential liability when Congress enacts a statute pursuant to its spending power, recipients of Title IX monies "must now assume that if conduct can be linked to sex discrimination – no matter how attenuated that link – this Court will impose liability under Title IX."[145] Although he may exaggerate the point, Justice Thomas's suggested ramifications are possible. The Court offers little direction in determining what behavior constitutes intentional discrimination; this may force funding recipients to be overly cautious when addressing allegations of sex discrimination. However, requiring federal funding recipients to tread carefully when dealing with a matter as delicate as sex discrimination may be a commendable outcome, rather than the deplorable outcome that Justice Thomas warned against.

In addition, Justice Thomas's contention that plaintiffs claiming retaliation are not required to prove the underlying discrimination does not undermine the majority's holding. The majority opinion asserts that retaliation for complaints of sex discrimination is itself a form of intentional sex discrimination.[146] Regardless of whether the underlying sex discrimination actually occurred, the retaliatory conduct constitutes an independent instance of actionable sex discrimination. Therefore, plaintiffs should not be required to prove the underlying sex discrimination in order to recover.

Justice Thomas's contention that petitioner does not fall within the statute's protected class is also without merit. The statute prohibits intentional sex discrimination, and, therefore, the Board should not be permitted to retaliate against the petitioner *or* the students he represented in making his complaints. Allowing the Board to escape punishment for its retaliatory

---

143. Mank, *supra* note 28, at 106.

144. *See supra*, notes 51-52 and accompanying text.

145. *Jackson*, 544 U.S. at 192 (Thomas, J., dissenting).

146. *Id.* at 173-74 (majority opinion) ("Retaliation is by definition an intentional act. . . . Moreover, retaliation is discrimination 'on the basis of sex' because it is an intentional response to the nature of the complaint: an allegation of sex discrimination.") (citations omitted)).

conduct simply because the person making the complaints was not the initial victim of discrimination is wholly illogical.

V. CONCLUSION

The decision in *Jackson v. Birmingham Board of Education* expands the scope of intentional discrimination prohibited by Title IX to include retaliation for complaints of sex discrimination, even when the complainant is not the victim of the underlying discrimination. While this expansive definition of intentional discrimination is logical in light of preceding case law interpreting what constitutes intentional discrimination, it is difficult to reconcile with the requirement that funding recipients receive "actual notice" of their potential liability under Spending Clause legislation. Federal funding recipients, therefore, must take note of this expansive definition and proceed cautiously when confronted with allegations of sex discrimination.

EMILY TUMBRINK BRACKSTONE

THE LICENSED PROFESSIONAL EXEMPTION IN CONSUMER PROTECTION: AT ODDS WITH ANTITRUST HISTORY AND PRECEDENT

MARK D. BAUER*

I. INTRODUCTION

Consumer protection law is more important today than ever. With an increasing number of jurisdictions creating caps and other limits on tort damages, particularly malpractice awards,[1] consumer protection may be the only recourse for injury brought on by the actions of licensed professionals, including doctors, lawyers, accountants, and architects.

Since consumer protection law originated in antitrust law, the study of consumer protection laws should begin with antitrust.[2] The same Federal Trade Commission ("FTC") Act[3] that was passed to protect consumers by supplementing the Sherman Act[4] was amended by Congress, in response to a limiting Supreme Court ruling, to clarify the FTC's authority to protect consumers from deceptive practices unrelated to competition between businesses. Today, the FTC is still entrusted with enforcement of both antitrust and consumer protection.[5] Consumer protection laws used in almost

---

* Assistant Professor of Law, Stetson University College of Law. B.A., The University of Chicago; J.D. Emory University. The author was formerly an attorney with the Federal Trade Commission's Bureau of Competition. The author wishes to thank: Dave O'Toole, Dee Pridgen, Joe Morrissey, and Spencer Weber Waller for their comments on an earlier draft of this Article; research assistant Jonathan Cohen, Stetson University College of Law J.D. Candidate 2006, for his help surveying the fifty states; Joe Paolucci, for his unique perspectives; Stetson's Faculty Support Services, particularly Silvia Manzanero; and CCH, Inc., especially John Arden and Bill Zale.

1. *See, e.g.*, James Dao, *A Push in States to Curb Malpractice Costs*, NY Times, Jan. 14, 2005, *available at* http://www.nytimes.com/2005/01/14/national/14malpractice.html?ex=1145592000&en=40cf68e98e390c5d&ei=5070 (last visited Apr. 19, 2006).

2. *See generally* Neil W. Averitt & Robert H. Lande, *Consumer Sovereignty: A Unified Theory of Antitrust and Consumer Protection Law*, 65 ANTITRUST L.J. 713, 744-45 (1997); Stephen Calkins, *FTC Unfairness: An Essay*, 46 WAYNE L. REV. 1935, 1935-37 (2000); Marshall A. Leaffer & Michael H. Lipson, *Consumer Actions Against Unfair or Deceptive Acts or Practices: The Private Uses of Federal Trade Commission Jurisprudence*, 48 GEO. WASH. L. REV. 521, 525 (1979-80).

3. Federal Trade Commission (FTC) Act, 15 U.S.C. §§ 41-58 (2000).

4. Sherman Act, 15 U.S.C. §§ 1-7 (2000).

5. FEDERAL TRADE COMMISSION, GUIDE TO THE FEDERAL TRADE COMMISSION, http://www.ftc.gov/bcp/conline/pubs/general/guidetoftc.htm (last visited Mar. 20, 2006).

every state in the nation were drafted borrowing language from the FTC Act—the same language the FTC uses today to enforce the antitrust laws.[6]

Courts in several states have recently considered whether consumer protection laws apply to the conduct of licensed professionals.[7] For years, the courts grappled with the question of whether professions constituted "trade or commerce" as used in the Sherman and FTC Acts. In 1975, the Supreme Court definitively answered the question: licensed professionals are subject to the antitrust laws.[8] Judges have ignored the common experience and ancestry of the antitrust and consumer protection laws and state legislative mandates to heed FTC precedent.[9] These decisions have created a complex web of conflicting decisions in consumer protection cases and are often based on highly subjective and speculative parameters.[10] Certainly, state legislatures may limit the scope of consumer protection laws;[11] however, judge-made consumer protection law that conflicts with the well-established antitrust law and FTC precedent has little support in history and constitutes poor public policy.

Consumer protection law is not an orphan. Although the actual wording of the consumer protection laws—the body of laws intended to protect consumers from unfair and deceptive acts and practices—shows its common roots with antitrust, it has rarely been considered a member of the antitrust family. It is unclear whether antitrust abandoned its progeny or whether consumer protection set off alone to find its own path. The solo course charted by consumer protection has led to anomalous and unfortunate jurisprudence.[12] As a relatively new field, with few common law decisions for guidance, judges in consumer protection suits look at far too many cases as ones of first impression; they often reach contradictory and questionable judgments on matters that were resolved in antitrust cases decades ago. Were the lineage of consumer protection laws unknown, such practices could be rationalized as the problems attendant to the adolescence of any new doctrinal area of law. But this is not the case.

Part II of this Article traces the common ancestry and development of the antitrust and consumer protection laws. A comprehensive study of the consumer protection laws in every state requires knowledge of the development of the Sherman and FTC Acts. Part III of this Article examines the treatment of licensed professionals under the antitrust and consumer

---

6. *E.g.*, CONN. GEN. STAT. ANN. § 42-110b (West 1958); *see* 15 U.S.C. § 45 (2000).

7. *See, e.g.*, Averill v. Cox, 761 A.2d 1083, 1086 (N.H. 2000).

8. Goldfarb v. Va. State Bar, 421 U.S. 773, 793 (1975) ("[H]olding that certain anticompetitive conduct by lawyers is within the reach of the Sherman Act . . . .").

9. *See, e.g.*, Rhode Island Unfair Trade Practices and Consumer Protection Act, R.I. GEN. LAWS § 6-13.1-3 (2001) ("[D]ue consideration and great weight" given to FTC precedent).

10. *See infra* Part IV.

11. *See, e.g.*, MD. CODE ANN., COM. LAW § 13-104 (LexisNexis 2005).

12. *See infra* Part IV.

protection laws. Recently, courts in several states have found that licensed professionals are exempt from suit under the consumer protection laws, contrary to the conclusion the Supreme Court reached for antitrust. Part IV argues that this development is not only contrary to public policy, but it is also based on unsound interpretations of the consumer protection laws. Part IV further suggests that in such matters when there is no relevant consumer protection case law, judges should turn to the common roots of consumer protection and antitrust; they should look to antitrust common law as authority.

II. THE DEVELOPMENT OF ANTITRUST AND CONSUMER PROTECTION LAWS

The familial connection between antitrust and consumer protection occurs on many levels. Many believe the purpose of antitrust is "consumer welfare," and therefore, it is a form of consumer protection.[13] The United States Federal Trade Commission ("FTC") mandate is to protect consumers, which it does through antitrust and consumer protection.[14] "Whether the FTC is combating telemarketing fraud, Internet scams, or price-fixing schemes, its primary mission is to protect consumers."[15] The most important connection between antitrust and consumer protection is Section 5 of the FTC Act; the FTC discharges both its antitrust and consumer protection missions through Section 5.[16] Understanding the relationship of Section 5 of the FTC Act to both antitrust and consumer protection, and indeed understanding how all state consumer protection laws were created, requires an understanding of the history of antitrust laws. The evolution and relationship of consumer protection to antitrust as traced through the rise fo the Sherman Act and the

---

13. *See, e.g.*, ROBERT H. BORK, THE ANTITRUST PARADOX: A POLICY AT WAR WITH ITSELF 57 (1978); Robert H. Lande, *Wealth Transfers as the Original and Primary Concern of Antitrust: The Efficiency Interpretation Challenged*, 34 HASTINGS L.J. 65, 83 (1982); *see also* HERBERT HOVENKAMP, FEDERAL ANTITRUST POLICY: THE LAW OF COMPETITION AND ITS PRACTICE 59 (3d ed. 2005).

14. *See generally* Federal Trade Commission Act, 15 U.S.C. §§ 41-58 (2000).

15. Message of the Chairman Timothy J. Muris, Federal Trade Commission: Performance & Accountability Report i (2003) *available at* http://www.ftc.gov/opp/gpra/2003pareport.pdf (last visited Mar. 20, 2006).

> The FTC works to ensure that the nation's markets are vigorous, efficient, and free of restrictions that harm consumers. Experience demonstrates that competition among firms yields products at the lowest prices, spurs innovation, and strengthens the economy. Markets also work best when consumers can make informed choices based on accurate information.
>
> To ensure the smooth operation of our free market system, the FTC enforces federal consumer protection laws that prevent fraud, deception, and unfair business practices. The FTC also enforces federal antitrust laws that prohibit anticompetitive mergers and other business practices that restrict competition and harm consumers.

Id.

16. 15 U.S.C. § 45 (2000).

creation of the FTC, followed by the evolution of the FTC Act and the creation of analogous state "Little FTC Acts" are given below.

A. The Rise of the Sherman Act

The passage of the Sherman Act in 1890 was the culmination of a political movement that sought to prohibit businesses, combined into trusts, from controlling price and output for entire industries.[17] Though a novel approach by government to regulate the economy, the Sherman Act merely reinforced an old idea, "that some economic practices are particularly suspect, somehow not normal, because they provide means, other than superior efficiency" for firms to gain competitive advantage.[18] In that respect, the antitrust laws and consumer protection laws serve the same general purpose.[19] Indeed, the Supreme Court observed that the Sherman Act was designed, in part, to yield "the lowest prices" and "the highest quality."[20]

The Sherman Act prohibits "[e]very contract, combination . . . , or conspiracy [which restrains] trade or commerce."[21] When the Supreme Court first considered a case brought under the Sherman Act, the court literally construed the term "every contract" to mean all contracts, without considering the reasonability of the restraint.[22] Just over ten years later, in *Standard Oil Co. of New Jersey v. United States*,[23] the Supreme Court re-evaluated its earlier interpretation of the Sherman Act and ruled that only "unreasonable" restraints of trade were prohibited.[24] The Court stated that Congress had not intended to prohibit "every contract" in restraint of trade, and instead that "judgment

---

17. *See* ANDREW I. GAVIL ET AL., ANTITRUST LAW IN PERSPECTIVE: CASES, CONCEPTS AND PROBLEMS IN COMPETITION POLICY 2 (2002) ("Trusts combined in one organization the power to make pricing and output decisions for entire industries, and .... [p]opular opposition to the trusts culminated in the 1890 passage of the 'Sherman Anti-trust Laws.'"); HANS B. THORELLI, THE FEDERAL ANTITRUST POLICY: ORIGINATION OF AN AMERICAN TRADITION 53 (1955) ("The state of the common law [regarding restraints on trade] at the end of the century demanded federal action.").

18. BORK, *supra* note 13, at 47.

19. *See generally* Averitt & Lande, *supra*, note 2, at 713 ("Antitrust and consumer protection law share a common purpose in that both are intended to facilitate the exercise of consumer sovereignty or effect consumer choice.").

20. N. Pac. Ry. Co. v. United States, 356 U.S. 1, 4 (1958).

21. 15 U.S.C. § 1 (2000).

22. United States v. Trans-Missouri Freight Ass'n, 166 U.S. 290, 341 (1897) ("[The Sherman Act] renders illegal *all* agreements which are a restraint of trade . . . ." (emphasis added)). *But see id.* at 371 (White, J., dissenting) (disagreeing with the majorities position that "even reasonable contract are embraced within the purview of the [Sherman] act"). The Supreme Court reaffirmed this decision in the following year. United States v. Joint Traffic Ass'n, 171 U.S. 505, 561-62, 573-78 (1898) (considering explicitly the question in *Freight Association*, whether *every* agreement means *all* agreements, for the third time).

23. 221 U.S. 1 (1911).

24. *See id.* at 59-60.

must in every case be called into play," leaving it to each judge to determine whether the Sherman Act had been violated.[25]

B. Creation of the Federal Trade Commission

Reactions to the *Standard Oil* decision were strong. Although John D. Rockefeller's Standard Oil trust was dissolved, Wall Street initially approved of the new "rule of reason" and stock prices rose.[26] Congress, which had passed a law that was clear and unambiguous, if overbroad, criticized the Supreme Court's decision as a judicial usurpation of legislative power.[27] "William Jennings Bryan, three-time Democratic Presidential nominee and [party activist], . . . declared 'The Trusts Have Won'[; f]ive senators . . . introduced bills" to restore the literal meaning of the Sherman Act.[28]

After the initial surprise of the *Standard Oil* decision, both businesses and the public wanted specific guidelines of precisely which activities were unacceptable.[29] Congress and business leaders were increasingly concerned

---

25. *Id.* at 63. Under the rubric of *Standard Oil*, a bifurcated analysis of potential Sherman Act violations arose and continues to this day; whereby, cases are analyzed either under the per se rule or the rule of reason. United States v. Trenton Potteries Co., 273 U.S. 392, 396-400 (1927). Under the per se rule, restraints on trade that are considered inherently anti-competitive are conclusively deemed unlawful. *Id.* "Practices considered per se illegal include: (1) [h]orizontal agreements . . . to fix prices . . . (2) [v]ertical agreements . . . to fix minimum resale prices . . . (3) [h]oritzontal territorial, customer, . . . and other market restraints between competitors . . . (4) [c]ertain competitively motivated group boycotts." WILLIAM C. HOLMES, ANTITRUST LAW HANDBOOK 182 (2005). For restraints of trade that are not inherently anticompetitive, a court will apply the rule of reason test. Under the rule of reason, a court considers and weighs procompetitive and anticompetitive factors to determine the impact, if any, upon competition. *See* Appalachian Coals, Inc. v. United States, 288 U.S. 344, 360-61 (1933); Bd. of Trade of Chi. v. United States, 246 U.S. 231, 238 (1918).

26. *Business Likes Oil Decision*, N.Y. TIMES, May 17, 1911, at 1.

27. S. REP. No. 62-1326, at 10-12 (1913); *see also* 47 CONG. REC. 1225 (1911) (statement of Sen. Newlands) (questioning whether the rule of reason would lead to consistent and fair interpretations of the Sherman Act).

28. Marc Winerman, *The Origins of the FTC: Concentration, Cooperation, Control, and Competition*, 71 ANTITRUST L.J. 1, 13 (2003) (quoting THE COMMONER, May 26, 1911, at 1).

29. *See* Gilbert H. Montague, *Anti-trust Laws and the Federal Trade Commission, 1914-1927*, 27 COLUM. L. REV. 650, 652 (1927) (discussing concerns of "political leaders, business leaders, lawyers and economist" that businessmen did not know what acts were legal or illegal under the "rule of reason" approach); *see also* Standard Oil. Co. v. United States, 221 U.S. 1, 102-03 (1911) (Harlan, J., concurring in part, and dissenting in part) ("But now, it is to be feared, we are to have, in cases without number, the constantly recurring inquiry—difficult to solve by proof—whether the particular contract, combination, or trust involved in each case is or is not an 'unreasonable' or 'undue' restraint on trade."); Jeffrey H. Liebling, *Judicial Usurpation of the F.T.C.'s Authority: A Return To The Rule of Reason*, 30 J. MARSHALL L. REV. 283, 289-90 (1996) ("Businessmen also feared that rather than providing specific guidelines designating what constituted an unreasonable restraint of trade, the Supreme Court's

that individual judges would deem on an *ad hoc* and *ex post facto* basis that certain behavior was presumptively illegal and other behavior was well-balanced against its potential harm.[30] A Bureau of Corporations was established within the Department of Commerce and Labor to buttress antitrust enforcement, but President Teddy Roosevelt was dissatisfied with the Bureau's limitations.[31] During the election of 1912, Roosevelt proposed a "commission" to protect consumers, shareholders, and workers.[32] Roosevelt went as far as to suggest that firms that voluntarily accepted the new commission's "regulation[s] and obeyed its orders in good faith would be shielded from antitrust prosecution."[33] Although Roosevelt lost the election of 1912, Roosevelt's proposed commission was created; it was created to President Woodrow Wilson's specifications and named the Federal Trade Commission.[34]

There was considerable debate and uncertainty as to what final form—if any—the new FTC would take. In a speech before Congress in January of 1914, President Wilson called for the creation of an investigatory commission, "only as an indispensable instrument of information and publicity."[35] Several months later, President Wilson proposed to the Senate that the commission's central role should be one of prosecution.[36]

While considering the FTC Act, Congress was also shoring up weaknesses in the Sherman Act with a series of provisions; the provisions became known as the "Clayton Act," named for Chairman Henry Clayton of the House Judiciary Committee.[37] The Clayton Act contains specific prohibitions on price discrimination, exclusive contracts, commodity tying arrangements, interlocking directorates, and holding companies.[38]

Debate continued on the best means for eliminating unfair competition. The concern remained whether the country would be better served by legislating a specific list of unlawful activities, the Clayton Act approach, or by granting extensive, but vague, powers of prosecuting "unfair competition"

Rule of Reason standard could never form a solid basis that businesses could follow in order to avoid violating the Sherman Act.").

30. *See* 47 CONG. REC. 1225 (1911) (statement of Sen. Newlands); Montague, *supra* note 29, at 652.

31. *See* Winerman, *supra* note 28, at 17-19.

32. *Id.* at 23.

33. *Id.* at 24.

34. *Id.* at 38.

35. WOODROW WILSON, TRUSTS AND MONOPOLIES, H.R. DOC. NO. 63-625, at 6 (2d Sess. 1914), *reprinted in* EARL W. KINTNER, THE LEGISLATIVE HISTORY OF THE FEDERAL ANTITRUST LAWS AND OTHER RELATED STATUTES, Part I, Vol. V at 3746 (1982) (presenting presidential speech before the joint session of Congress on January 20, 1914).

36. *See* Winerman, *supra* note 28, at 62; *see* NEWLANDS, FEDERAL TRADE COMMISSION, S. REP. NO. 63-597, at 7 (2d Sess.1914).

37. *See* Winerman, *supra* note 28, at 53-55.

38. Clayton Act, 15 U.S.C. §§ 12-27 (2000).

to the FTC.[39] The final bill creating the Federal Trade Commission adopted the second option and afforded the FTC enormous and vague power to police "unfair competition."[40] Section 5 of the FTC Act, which granted this awesome power, received opposition from both critics of this broad grant of discretion and those who merely distrusted the proposed agency.[41]

Section 5's proponents came from two fields of law: antitrust and business torts.[42] The antitrust proponents saw *Standard Oil* as having engaged in "unfair competition" as well as monopolization; they believed the broad powers extended to the FTC would give the government an additional weapon in its arsenal to combat the next Standard Oil Company.[43] The proponents concerned with business torts were seeking an expansion of traditional business torts and a means to reach comparatively new forms of unfair conduct.

Originally "unfair competition" consisted chiefly of the expressed or implied representation that the goods or business of one man are the goods or business of another. Now, however, it is given a much broader meaning, and is not limited to mere infringements of trademarks . . . yet designed to deceive the public . . . .[44]

Advocates of Section 5 in the Senate supported their argument by extensively quoting Harry D. Nim's seminal work on unfair business competition:

> "Unfair competition is not confined to acts directed against the owners of trade-marks or trade names, but exists wherever unfair means are used in trade rivalry. Equity looks not at what business the parties before the court are engaged in, but at the honesty or dishonesty of their acts. It is unfair to pass off one's goods as those of another person; it is unfair to imitate a rival's trade name or label; but he who seeks to win trade by fair means or foul is not limited to these methods. He may copy and imitate the actual goods made or sold by a competitor; he may libel or slander these goods, make fraudulent use of a family name, of trade secrets, of corporate names, of signs, of threats of action; he may construct buildings which are reproductions of peculiar buildings of a rival, thus producing confusion in the minds of purchasers . . . ."[45]

---

39. *See* Winerman, *supra* note 28, at 48, 57-58, 62. *See generally* 15 U.S.C. §45 (2000).

40. *See* 15 U.S.C. § 45, also known by its internal numbering as FTC Act § 5.

41. Winerman, *supra* note 28, at 69-74.

42. *Id.* at 75-76.

43. *Id.* at 76.

44. 51 CONG. REC. 11,228 (1914) (statement of Sen. Robinson), *reprinted in* KINTNER, *supra* note 35, at 3997.

45. *Id.* (quoting HARRY D. NIMS, LAW OF UNFAIR BUSINESS COMPETITION 1 (1st ed. 1909)) *reprinted in* KINTNER, *supra* note 35, at 3997. The art of deceiving the consumer into thinking a product was actually created by another is called "palming off." *Id.*

It appears from this debate that at least some members of the Senate were concerned at the inception of the FTC Act that certain types of unfair competition created deception and "confusion in the minds" of consumer.[46]

In fact, this approach was entirely consistent with granting the FTC purview over consumer unfairness, or, in more modern terms, consumer protection. Unfair consumer practices are a form of unfair competition; by engaging in unfair consumer practices, a firm might "realize sales and profits not otherwise obtainable, and could thereby gain an improper competitive advantage over other firms."[47]

This was not a lapse or loophole on the part of Congress. The entire concept of "consumer protection" was so novel that the 63rd Congress used words archaic to the modern doctrine. The very beginnings of a consumer movement had been brewing for some years. In 1848, Congress passed the Import Drugs Act to restrict the sale of "counterfeit, contaminated, diluted, and decomposed drugs" in the United States.[48] "Starting in the 1860s, scientists, physicians, and pharmacists agitated for regulation of foods and drugs by local boards of health. In some instances, businesses supported regulation, because it was difficult for the honest producer to compete with someone who sold adulterated products."[49] In 1862 the Department of Agriculture established a laboratory to analyze food, fertilizers, and other agricultural substances.[50] In 1865, a federal law was passed to prohibit the importation of diseased cattle and pigs.[51] In 1906, Upton Sinclair published his consumerist classic *The Jungle*,[52] and later that year Congress passed the Pure Food and Drug Act and the Meat Inspection Act.[53]

The two principal sponsors of the FTC Act in the Senate clearly intended Section 5 to endow the FTC with powers attendant to modern consumer protection. Senator Albert Baird Cummins of Iowa urged the Senate to adopt Section 5 by describing a situation in which a firm adopted a product name

---

46. *See id.*

47. Neil W. Averitt, *The Meaning of "Unfair Acts or Practices" in Section 5 of the Federal Trade Commission Act*, 70 GEO. L.J. 225, 230-31 (1981-1982).

48. ROBERT N. MAYER, THE CONSUMER MOVEMENT 16 (1989).

49. *Id.* (citation omitted).

50. *Id.*

51. *Id.*

52. UPTON SINCLAIR, THE JUNGLE (1906).

53. Federal Meat Inspection Act, ch. 2907, 34 Stat. 1260 (1907) (current version at 21 U.S.C. §§ 601 - 686 (2000)); Pure Food and Drug Act, ch. 3915, 34 Stat. 768 (1906) (superceded by Food, Drug and Cosmetic Act of 1938); MAYER *supra* note 48, at 17.

> Sinclair was far less concerned about the unsanitary conditions of the meat-packing industry than he was about the inhumane conditions under which the industry's employees worked and lived. . . . To Sinclair's disappointment, the public outrage his book inspired was channeled into consumer protection reform. He is reported to have said, "I aimed for the nation's heart, but I hit its stomach instead."

Id. at 17-18.

that was confusingly similar to that of a popular, established brand.[54] Section 5 would give the FTC power to protect consumers from such deception. Senator Francis Griffith Newlands of Nevada, arguing in favor of Section 5, suggested that the FTC should be able to declare that false claims regarding the quality of a product are unlawful.[55]

Supporters in the House of Representatives echoed the sentiment of protecting consumers. Representative Stevens of Minnesota urged that "fraud . . . is the basis and essence of unfair competition."[56] He further urged that the Senate's wording of "unfair competition" be changed to "unfair methods of competition" in order to broaden the scope of conduct termed fraudulent.[57] The Senate's emphasis on prosecution and broad powers over unfair competition prevailed in the final bill, but the House was successful in changing "unfair competition" to "unfair methods of competition."[58] President Wilson signed the bill creating the FTC, and, less than one month later, he signed the Clayton Act.[59] According to the Senate Committee on Interstate Commerce, Section 5 of the FTC Act, which "declares unfair competition in commerce to be unlawful," was "[o]ne of the most important provisions of the bill"[60]

Having tried and failed to delineate all forms of unfair competition, Congress left it to the new FTC to determine the definition of "unfair" under its Section 5 power.[61] The connection between antitrust law and consumer protection was now clear and enshrined in law. Congress's design for the FTC's mission unified antitrust with consumer protection and encompassed everything from [price] predation to fraud. Traditional consumer protection violations, such as deception, were reached by this concept on the theory that they diverted trade from honest firms and ultimately harmed competition. Congress's language thus did not appear to state two separate and independent

---

54. 51 CONG. REC. 11,105 (1914) (statement of Sen. Cummings), *reprinted in* KINTNER, *supra* note 35, at 3952.

55. *See id.* at 11,109 (statement of Sen. Newlands), reprinted in KINTNER, *supra* note 35, at 3960.

56. *Id.* at 14,936 (statement of Rep. Stevens) *reprinted in* KINTNER, *supra* note 35, at 4742.

57. *See id.* at 14,937 *reprinted in* KINTNER, *supra* note 35, at 4743.

58. Winerman, *supra* note 28, at 90; *see* 15 U.S.C. §§ 41-58 (2000); H.R. REP. NO. 63-1142 at 19 (1914) (Conf. Rep.).

59. Clayton Act, ch. 323, 38 Stat. 730 (1914) (codified as amended at 15 U.S.C. §§ 12-27 (2000)); Federal Trade Commission Act, ch. 311, 38 Stat. 717 (1914) (codified as amended at 15 U.S.C. §§ 41-58 (2000)).

60. Report of the Senate Committee on Interstate Commerce, S. REP. 63-597, at 13 (1914).

61. *Id.* The Senate Committee on Interstate Commerce came up with twenty forms of unfair competition before abandoning its effort to itemize all forms and leaving it to the fledgling FTC "to determine what practices were unfair." *Id.*

goals for the FTC, but rather to visualize areas of activity that the legislature considered to be closely related.[62]

C. *Evolution of "Unfair Methods of Competition" and the Relationship of Section 5 of the FTC Act to the Sherman Act*

Early on, the broad sweep of Section 5 of the FTC Act was met by judicial inconsistency—not dissimilar to the Supreme Court's initial, inconsistent interpretations of the Sherman Act.[63] In 1920, in a purely antitrust matter brought by the FTC under Section 5, the Supreme Court held that '[t]he words 'unfair method of competition' are not defined by the statute and their exact meaning is in dispute. It is for the courts, not the commission, ultimately to determine as matter [sic] of law what they include."[64] Justice Brandeis dissented, arguing that it was impossible to itemize every form of unfair competition and that Congress had left it to the FTC to determine whether a specific method of competition in a particular case was unfair.[65] Two years later, in a purely consumer protection case, the Supreme Court found a Section 5 violation because it resulted in at least an indirect injury to competition.[66]

The clear wording of Section 5, targeting "unfair methods of competition," required the FTC to show some harm to competition in all consumer protection cases. The tensions came to a head in 1931 in the case *FTC v. Raladam, Co.*[67]

---

62. Averitt & Lande, *supra* note 2, at 745 (footnotes omitted); *see, e.g.*, FTC v. R.F. Keppel & Bro., 291 U.S. 304, 313 (1934) (combining damages arising from competition and fraud out of same practice); FTC v. Winsted Hosiery Co., 258 U.S. 483, 494 (1922).

> The honest manufacturer's business may suffer, not merely through a competitor's deceiving his direct customer, the retailer, but also through the competitor's putting into the hands of the retailer an unlawful instrument, which enables the retailer to increase his own sales of the dishonest goods, thereby lessening the market for the honest product.

Winsted Hosiery, 258 U.S. at 494.

63. *Compare* United States v. Trans-Missouri Freight Ass'n, 166 U.S. 290, 341 (1897) (holding that all contracts that restrain trade are under the purview of the Sherman Act); United States v. Joint Traffic Ass'n, 171 U.S. 505, 568-69 (1898) (stating that Congress has the power to determine that all contracts that restrain interstate commerce are invalid), *with* Standard Oil Co. v. United States, 221 U.S. 1, 51, 60 (1911) (holding that contracts with partial restraints that are otherwise considered reasonable at common law are valid); Bd. of Trade of Chi. v. United States 246 U.S. 231, 241 (1918) ("Every board of trade and nearly every trade organization imposes some restraint upon the conduct of business by its members."). These cases show the evolution of Supreme Court's interpretation of Section 1 of the Sherman Act from interpreting the word "every" as "all" to the bifurcated evidentiary rules of the per se rule and rule of reason.

64. FTC v. Gratz, 253 U.S. 421, 427 (1920).

65. *Id.* at 436 (Brandeis, J., dissenting).

66. *Winsted Hosiery*, 258 U.S. at 491-94 (finding the underwear producer's practice, implying that products were all-wool when instead the products were combined with cotton, constitutes an unfair method of competition).

67. 283 U.S. 643 (1931).

In *Raladam*, a patent medicine manufacturer marketed an "obesity cure" that contained "desiccated thyroid" and could not be sold safely without the supervision of a physician.[68] Raladam represented to the public "that the preparation [was] the result of scientific research, knowledge and accuracy, that it [was] safe and effective and may be used without discomfort inconvenience, or danger of harmful results to health."[69] The FTC found the manufacturer's actions to be misleading, unfair, and deceptive and issued a cease and desist order.[70] In a unanimous decision, the Supreme Court held that the FTC was without jurisdiction to prohibit false advertising without proof that the advertisement adversely affected "present or potential competitors."[71]

The burden placed on the FTC was not insurmountable; arguably, if the FTC produced one complaining competitor at a hearing, it would have met the *Raladam* burden.[72] Congress, however, reacted to *Raladam* by beginning work to amend the FTC Act.[73]

The result was the Wheeler-Lea Amendment to the FTC Act, passed in 1938.[74] Congress added language to Section 5 declaring unlawful not only "unfair methods of competition," but also "unfair or deceptive acts or practices."[75] Congress was clear, however, that this was not an expansion of the FTC's powers, but a procedural change made to comply with a Supreme Court order.[76] Senator Wheeler, one of the bill's primary sponsors, stated that the "bill is a reenactment of the present law . . . not for the purpose of adding to [the FTC's] powers but for the purpose of aiding them in carrying out their

---

68. *Id.* at 644-45. "Findings supported by the evidence, warrant the conclusion that the preparation is one which cannot be used generally with safety to physical health except under medical direction and advice." *Id.* at 646.

69. *Id.* at 645.

70. *Id.* at 645-46.

71. *Id.* at 648-49. "If the necessity of protecting the public against dangerously misleading advertisements of a remedy sold in interstate commerce were all that is necessary to give the Commission jurisdiction, the [cease and desist] order could not successfully be assailed. But this is not all." *Id.* at 646.

72. Averitt, *supra* note 47, at 233.

73. *Id.*; *see* H.R. REP. NO. 75-1613, at 3 (1937) ("Since it is the purpose of Congress to protect the consumer as well as the honest competitor, the Commission should be empowered to prevent the use of unfair or deceptive acts or practices in commerce, regardless of whether such acts or practices injurously affect a competitor."); S. REP. NO. 74-1705, at 5 (1936) ("Without these amendatory words, there is a question whether the Commission has the jurisdiction of a case of deceptive or other unfair practices where it develops that the offender has no competitor . . . or that all competitors in an industry are equally guilty."); EARL W. KINTNER, A PRIMER ON THE LAW OF DECEPTIVE PRACTICES 58-59 (2d ed. 1978) ("The *Raladam* decision struck a hard blow . . . . Congressional momentum soon developed to broaden the FTC's powers to enable the Commission to protect consumers as well as honest competitors.")

74. Federal Trade Commission Act, ch. 49, 52 Stat. 111 (1938).

75. *Id.* (codified as amended at 15 U.S.C. § 45(a)(1) (2000)); Averitt, *supra* note 47, at 234.

76. Averitt, *supra* note 47, at 234.

present powers. . . . It does not change the fundamentals of the act in the slightest."[77] Congressman Lea, the primary sponsor in the House, believed the amendment was "procedural."[78]

Furthermore, Wheeler-Lea made it clear that the FTC does not have two separate and independent goals—antitrust and consumer protection—but one area of closely related activity.[79] Section 5 of the FTC Act, at its most expansive, allows the FTC to "fill the interstices of the Sherman Act."[80] In other words, the most expansive reading of Section 5 permits an "occasional case" to enforce well-established public policies, which support antitrust and consumer protection, not specifically itemized in the Sherman (or Clayton) Act.[81]

In 1972, in *FTC v. Sperry & Hutchinson Co.*,[82] the Supreme Court suggested in dicta that the scope of Section 5 of the FTC Act extended well beyond the reach of the Sherman Act (as well as beyond the Clayton Act).[83] The *Sperry & Hutchinson* decision was very controversial because it suggested that a cause of action might rest on some theory outside the Sherman and Clayton Acts, without any frame of reference, and without any guidance to either businesses or the FTC.[84] To date, lower courts have not followed the Supreme Court's dicta in expanding the scope of the FTC Act.[85]

---

77. 80 CONG. REC. 6,589 (1936) (statement of Sen. Wheeler).

78. *See* 83 CONG. REC. 391 (1938) (statement of Rep. Lea).

79. Averitt, *supra* note 47, at 281.

80. *Id.* at 282; *see* 51 CONG. REC. 11,236 (1914) (statement of Sen. Cummings) ("[T]he only purpose of section 5 [is] to make some things punishable, to prevent some things, that can not [sic] be punished or prevented under the antitrust law.").

81. Averitt, *supra* note 47, at 282; *see, e.g.*, Grand Union v. FTC, 300 F.2d 92, 98-99 (2d Cir. 1962) (noting that Section 5 allows the FTC to ban practices at odds with the primary purposes of the Sherman and Clayton Acts); *see also* Caswell O. Hobbs, *Antitrust and Consumer Protection—Exploring The Common Ground* Prepared Remarks for the Panel Discussion on "Under One Unbrella [sic]: Integrating the Competition and Consumer Protection Missions" at the Federal Trade Commission, 90th Anniversary Syposium [sic] (Sept. 22, 2004), http://www.ftc.gov/ftc/history/docs/041123hobbs.pdf#search='hobbs%20antitrust' (last visited Mar. 20, 2006) (commenting on the FTCs ability to reach "beyond the bounds of the Sherman and Clayton Acts").

82. 405 U.S. 233 (1972).

83. *See id.* at 244 ("[U]nfair competitive practices were not limited to those likely to have anticompetitive consequences after the manner of the antitrust laws . . . . Congress, through § 5, charged the FTC with protecting consumers as well as competitors.").

84. HOLMES, *supra* note 25, at 748-49.

85. The FTC itself has been cautious in seeking to expand the definition of "unfair" in Section 5 of the FTC Act. Commissioner Thomas B. Leary (writing as an individual and not speaking for the FTC) noted that the FTC definition has changed from one emphasizing "moral and ethical concepts" to one giving "primacy to economic factors, and introduc[ing] the notion of consumer responsibility." Thomas B. Leary, *Unfairness and the Internet*, 46 WAYNE L. REV. 1711, 1713 (2000). This may be in large part due to Congress's choice to amend the FTC Act, perhaps to restrict the meaning of unfair, by inserting the following language into the Act:

In fact, the United States Court of Appeals for the Second Circuit considered this issue in the context of unilateral bad acts—the so-called gap between Section 1 and Section 2 of the Sherman Act—when a single firm's conduct is in question, but the firm is not alleged to have monopolized or attempted to monopolize.[86] The Second Circuit held that the FTC's power under Section 5, beyond the Sherman and Clayton Acts, is generally limited to matters that violate the "spirit" of those Acts.[87] The court declined to "break new ground" and extend Section 5 beyond already established matters of collusive, coercive, predatory, restrictive or deceitful conduct. . . . A test based solely upon restraint of competition, even if qualified by the requirement that the conduct be 'analogous' to an antitrust violation, is so vague as to permit arbitrary or undue government interference with the reasonable freedom of action that has marked our country's competitive system.[88]

So while there may be room—at least in unusual situations—for a more expansive view of Section 5 of the FTC Act, it is more likely that courts will interpret Section 5 more narrowly,[89] defined by the limits of the Sherman and Clayton Acts.[90]

> [whether] the act or practice causes or is likely to cause substantial injury to consumers which is not reasonably avoidable by consumers themselves and not outweighed by countervailing benefits to consumers or to competition . . . . [T]he Commission may consider established public policies as evidence . . . . [but] public policy considerations may not serve as a primary basis for such determination.

Id. at 1712 (alteration in original) (quoting 15 U.S.C. § 45(n) (2000)).

86. *See* E.I. Du Pont De Nemours & Co. v. FTC, 729 F.2d 128, 133 (2d Cir. 1984) ("The complaint attacked . . . noncollusive practices . . . [and] did not complain that the practices . . . had been undertaken for other than legitimate business purposes."). *See generally* Berkey Photo, Inc. v. Eastman Kodak Co., 603 F.2d 263, 294 (2d Cir. 1979), *cert. denied* 444 U.S. 1093 (1980).

87. *See id.* at 136-37 ("[The FTC] may bar incipient violations of [the Sherman and Clayton Acts], and conduct which, although not a violation of the letter of the antitrust laws, is close to a violation or is contrary to their spirit." (citations ommitted)).

88. *Id.* at 137.

89. *See, e.g.*, Official Airline Guides, Inc. v. FTC, 630 F.2d 920, 927 (2d Cir. 1980) (limiting interpretation of Section 5 to the Sherman Act, rather than the "social, political or personal"); Boise Cascade Corp. v. FTC, 637 F.2d 573, 581 (9th Cir. 1980) (stating that, absent a Sherman Act violation, the FTC must prove anticompetitive effects).

90. HOLMES, *supra* note 25, at 748; *see, e.g.*, FTC v. Ind. Fed'n of Dentists, 476 U.S. 447, 451-52, 455 (1986) (using Sherman Act violations to establish violation of Section 5 of the FTC Act); *see also* Advertising of Ophthalmic Goods and Services, 43 Fed. Reg. 23, 992, 24,003 (Jun. 2, 1978) (codified at 16 C.F.R. Pt. 456) (opining that many practices that violate Section 5 are simultaneously antitrust and consumer protection violations). Interpreting FTC Act Section 5 jurisprudence is not for the fainthearted. Without careful recognition and analysis of its historical underpinnings, it would be possible for a single unlawful act with both antitrust and consumer protection ramifications to be analyzed using the identical statutory language of Section 5 and citing the same case law, but reaching opposite conclusions as to what the statute and case law actually mean. Perhaps to avoid this dilemma, the FTC "does not treat the

D. Relationship of Section 5 of the FTC Act to State Consumer Protection Laws

Although Congress defined the FTC's mission broadly, the FTC has taken a narrow path in choosing its enforcement actions—likely because of budgetary constraints.[91] In the 1960s, both the American Bar Association and a Ralph Nader group criticized the FTC for "wasting its time on trivial cases while overlooking more serious forms of consumer fraud" and for "long delays in the few consumer protection cases that were brought."[92] Perhaps because of a real or perceived lack of interest in Washington, states began enacting their own consumer protection statutes.[93] Practitioners refer to these statutes as "Little FTC Acts" because they contain identical language to the FTC Act, typically forbidding "unfair competition and unfair deceptive acts and practices."[94] By the early 1980s, all fifty states and the District of Columbia had enacted one or more consumer protection laws.[95]

States adopted their own versions of the FTC Act for reasons in addition to a decline in federal actions to protect consumers. In particular, consumers found it too difficult to prevail in state common law actions for fraud and warranty;[96] the disparity in bargaining power between consumer and merchant led to inequalities under state law;[97] the costs for individual consumers to litigate fraud and warranty disputes in state courts were often prohibitive;[98] and neither the federal government, nor the states could successfully bring all necessary and appropriate consumer litigation—individuals needed to be empowered as private attorneys general, complete with the ability to recover costs, attorney fees, and multiple damages.[99]

Despite the moniker of "Little FTC Acts," state consumer protection laws vary considerably. Some states have chosen to copy Section 5 of the FTC Act

---

theoretical division between its two bureaus [competition and consumer protection] as if it were entirely watertight." Averitt, *supra* note 47, at 283-84.

91. CCH STATE UNFAIR TRADE PRACTICES LAW ¶1000 (Mark D. Bauer ed. CCH 2000).

92. DEE PRIDGEN, CONSUMER PROTECTION AND THE LAW § 8:2 (2006 ed.).

93. CCH, *supra* note 91, at ¶ 1000. This resurgence in the consumer protection movement may also be traced to a speech made by President John F. Kennedy when he set forth four basic rights of consumers, including a right to be protected against "fraudulent, deceitful or grossly misleading" conduct. 108 CONG. REC. 4,168 (1962) (message from Pres. Kennedy).

94. *See* PRIDGEN, *supra* note 92, § 3:5; CCH, *supra* note 91, at ¶¶ 1050, 1200. While there is some variety between these state laws, all will be referred to generically as "Little FTC Acts."

95. CCH, *supra* note 91, at ¶ 1000.

96. PRIDGEN, *supra* note 92, § 3:2.

97. *Id.*

98. *Id.*

99. *See id.* "Private attorneys general" permits recovery by consumers. Ill. Brick Co. v. Illinois, 431 U.S. 720, 764 n.23 (1977) (Brennan, J., dissenting).

in its entirety; some states have adopted all or part of three model state consumer protection laws (plus official suggested variations); some states have copied the FTC or a model act but changed some of the wording; and some states have combined two or more of these approaches.[100] Several states have copied the statutes of other states, and many states have more than one consumer protection act.[101] Although the states have varied approaches to protecting their consumers, there are significant commonalities among most states, and virtually all of these state laws can be traced directly to Section 5 of the FTC Act.

The Unfair Trade Practices and Consumer Protection Law was developed as a model state consumer protection law by the FTC in conjunction with the Committee on Suggested State Legislation of the Council of State Governments.[102] The law's central prohibition is against "unfair methods of competition and unfair or deceptive acts or practices" and its language is taken directly from Section 5 of the FTC Act.[103] Twenty-five states have adopted language from this model act,[104] and at least three other states have enacted language identical to Section 5 of the FTC Act.[105] In sum, twenty-eight states have prohibited "unfair methods of competition and unfair, deceptive acts or practices" by adopting language from Section 5 of the FTC Act.

A second model act, the Uniform Deceptive Trade Practices Act was drafted by the National Conference of Commissioners on Uniform State Laws in 1964 and rewritten in 1966.[106] Limiting private actions to injunctive relief, the model act was intended only as a remedy for business competitors.[107] Most

---

100. PRIDGEN, *supra* note 92, §§ 3:4-:7.

101. For example, "[t]he Iowa Consumer Fraud Act was in part patterned after the Illinois Consumer Fraud Act." JONATHAN SHELDON & CAROLYN L. CARTER, UNFAIR AND DECEPTIVE ACTS AND PRACTICES § 3.4.2.5, at 107 n.168 (5th ed. 2001); *see also* State *ex. rel.* Miller v. Hydro Mag, Ltd., 436 N.W.2d 617, 621 (Iowa 1989).

102. SHELDON & CARTER, *supra* note 101, § 3.4.2.2, at 106.; *see also* Restatement (Third) of Unfair Competition §1 at 13 (1993); COMMITTEE OF STATE OFFICIALS ON SUGGESTED STATE LEGISLATION, THE COUNCIL OF STATE GOVERNMENTS, SUGGESTED STATE LEGISLATION C-4 (1969) [hereinafter SUGGESTED STATE LEGISLATION].

103. RESTATEMENT (THIRD) OF UNFAIR COMPETITION § 1, at 13 (1993); SUGGESTED STATE LEGISLATION, *supra* note 102, at C-4.

104. The states are Alaska, Connecticut, Florida, Georgia, Hawaii, Idaho, Illinois, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Mississippi, Montana, Nebraska, New Hampshire, North Carolina, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Vermont, Washington, and West Virginia. SHELDON & CARTER, *supra* note 101, § 3.4.2.2, at 106 nn.153-54 & 156.

105. The states are California, Utah, and Wisconsin. *Id.* at 106 nn. 157-58. "Utah has a statute prohibiting 'unfair methods of competition' dating from 1937, thus antedating even the 1938 Wheeler Lea Amendments to the FTC Act that added the prohibition against unfair or deceptive practices." *Id.* at 106.

106. *Id.* § 3.4.2.4 at 107.

107. PRIDGEN, *supra* note 92, § 3:4.

of the thirteen states that adopted this model have added a second statute that provides additional remedies for consumers.[108]

The National Conference of Commissioners on Uniform State Laws and the American Bar Association drafted and offered to the states a third model act, the Uniform Consumer Sales Practices Act.[109] The Uniform Consumer Sales Practices Act is narrower than the other model acts; it applies only to "consumer transactions."[110] Only three states have adopted this model act.[111]

Fourteen states and the District of Columbia, often by incorporating the three model acts or by borrowing from a kindred state, have each developed a unique form of consumer protection act.[112] These acts borrow heavily from both the FTC and the model acts.[113]

While the fifty-one Little FTC Acts are similar to the federal act, each (with the exception of Iowa's Consumer Fraud Act) differs from the federal version in one important way: the Little FTC Acts generally provide for private enforcement and private remedies, and about half provide for treble damages akin to the antitrust laws.[114]

Along with the drafters of these model rules, state legislators were evidently wary of the dangers associated with each state passing laws similar to the FTC Act while their respective courts provided dissimilar interpretations. To combat this potential problem, the majority of these statutes urge harmonization with FTC precedent; some even go so far as to mandate that courts be guided by the case law interpreting similar statutes of other states.[115] State courts often show great deference to FTC actions even when the statute does not explicitly refer to the precedential value of FTC decisions.[116] Even

---

108. *Id.* at 120. The thirteen states that have passed this act are Colorado, Delaware, Georgia, Hawaii, Illinois, Maine, Minnesota, Nebraska, Nevada, New Mexico, Ohio, Oklahoma and Oregon. *Id.*

109. SHELDON & CARTER, *supra* note 101, § 3.4.2.3, at 106.

110. *Id.* § 3.4.2.3; John A. Sebert, Jr., *Enforcement of State Deceptive Trade Practices Statutes*, 42 Tenn. L. Rev. 689, 701 (1975).

111. SHELDON & CARTER, *supra* note 101, § 3.4.2.3, at 106. The states are Kansas, Ohio and Utah. *Id.* at 106 n.162

112. *Id.* § 3.4.2.5, at 107.

113. *Id.* The states are Arizona, Arkansas, Delaware, Illinois, Indiana, Iowa, Michigan, Missouri, New Jersey, New York, North Dakota, South Dakota, Virginia, Wyoming and the District of Columbia. *Id.* at 107 nn.168-69

114. *Id.* §8.2, at 608-09; *see id.* at § 8.4.2.1, at ***621-22; *see also* IOWA CODE ANN. §§ 714.16 to 714.16A (West 2003).

115. *See* Jack E. Karns, *State Regulation of Deceptive Trade Practices Under "Little FTC Acts": Should Federal Standards Control?* 94 DICK. L. REV. 373, 379 (1990).

116. *See, e.g.*, People *ex rel.* Mosk v. Nat'l Research Co., 20 Cal. Rptr. 516, 521-22 (Cal. Ct. App. 1962); Guste v. Demars, 330 So. 2d 123, 125 (La. Ct. App. 1976); State *ex rel.* Lefkowitz v. Colo. State Christian Coll., Inc., 346 N.Y.S.2d 482, 487 (Sup. Ct. 1973); Marshall v. Miller, 276 S.E.2d 397, 399 (N.C. 1981); Wesware, Inc. v. State, 488 S.W.2d 844, 848 (Tex. App. 1972). Furthermore, courts have made no distinction as to the fact that the FTC Act is primarily enforced at the administrative level while state Little FTC Acts are enforced by

when a Little FTC Act makes no specific reference to the federal FTC Act, state courts will, nevertheless, look to federal law for guidance due to the similarities in language and because the FTC Act is considered a "rational source of authority on the meaning of the broad prohibitions" in the Little FTC Acts.[117] Table One below summarizes each state's required degree of harmonization with the FTC Act and the similar Little FTC Acts of other states.

courts. *See generally* People *ex rel.* Dunbar v. Gym of Am., Inc., 493 P.2d 660, 668-69 (Colo. 1972). At least one court has given weight to FTC advisory opinions in making rulings. PMP Assocs., Inc. v. Globe Newspaper Co., 321 N.E.2d 915, 919 (Mass. 1975).

117. Leaffer & Lipson, *supra* note 2, at 534.

Table 1. Harmonization of the FTC Act and Little FTC Act Precedents

STATE	Harmonization with FTC and/or other states with similar Little FTC Acts?
Alabama	"[D]ue consideration and great weight" given to FTC precedent[118]
Alaska	"[D]ue consideration and great weight" given to FTC precedent[119]
Arizona	FTC precedent is a guide[120]
Arkansas	Not stated
California	Not stated
Colorado	Not stated
Connecticut	Guided by FTC precedent[121]
Delaware	Not stated
District of Columbia	Not stated
Florida	To be construed consistently with FTC precedent[122]
Georgia	Uniform Deceptive Trade Practices Act: Courts shall construe Act uniformly with other states enacting it;[123] Fair Business Practices Act: To be construed consistently with FTC precedent[124]
Hawaii	To "be construed in accordance with . . . similar federal antitrust statutes"[125]
Idaho	Construed uniformly with FTC precedent[126]

118. ALA. CODE § 8-19-6 (LexisNexis 2002).
119. ALASKA STAT. § 45.50.545 (2004).
120. ARIZ. REV. STAT. ANN. § 44-1522 (2002).
121. CONN. GEN. STAT. ANN. § 42-110b (West 2000).
122. FLA. STAT. ANN. § 501.202 (West 2002).
123. GA. CODE ANN. § 10-1-375 (2000).
124. *Id.* at § 10-1-391.
125. HAW. REV. STAT. ANN. § 480-3 (1985).
126. IDAHO CODE ANN. § 48-618 (2003).

Illinois	Consumer Fraud and Deceptive Business Practices Act: Consideration to be given to FTC precedent;[127] Uniform Deceptive Trade Practices Act: To be construed in a manner uniform to other states that have passed the Act.[128]
Indiana	Not stated
Iowa	Not stated
Kansas	Not stated
Kentucky	Not stated
Louisiana	Not stated
Maine	Unfair Trade Practices Act: Courts are to be guided by FTC precedent;[129] Uniform Deceptive Trade Practices Act: Courts shall construe Act uniformly with other states enacting it[130]
Maryland	"[D]ue consideration and weight" given to FTC precedent[131]
Massachusetts	Courts are to be guided by FTC precedent[132]
Michigan	Not stated
Minnesota	Uniform Deceptive Trade Practices Act: Courts shall construe Act uniformly with other states enacting it[133]
Mississippi	Courts are to be guided by FTC precedent[134]
Missouri	Not stated

127. 815 ILL. COMP. STAT. ANN. 505/2 (West 1999).
128. *Id.* at 510/5.
129. ME. REV. STAT. ANN. tit. 5, § 207(1) (2002).
130. ME. REV. STAT. ANN tit. 10, § 1215 (1997).
131. MD. CODE ANN., COM. LAW § 13-105 (LexisNexis 2005).
132. MASS. GEN. LAWS ANN. ch. 93A, § 2(b) (West 1997).
133. MINN. STAT. ANN. § 325D.47 (West 2004).
134. MISS. CODE ANN. § 75-24-3 (2000).

Montana	"[D]ue consideration and great weight" given to FTC precedent[135]
Nebraska	Uniform Deceptive Trade Practices Act: Courts shall construe Act uniformly with other states enacting it[136]
Nevada	Not stated
New Hampshire	Courts to be guided by FTC precedent[137]
New Jersey	Not stated
New Mexico	Courts are to be guided by FTC precedent[138]
New York	Not stated
North Carolina	Not stated
North Dakota	Not stated
Ohio	Consumer Sales Practices Act: "[D]ue consideration and great weight" given to FTC precedent[139]
Oklahoma	Not stated
Oregon	Not stated
Pennsylvania	Not stated
Rhode Island	"[D]ue consideration and great" weight given to FTC precedent[140]
South Carolina	Courts are to be guided by FTC precedent[141]
South Dakota	Not stated

135. MONT. CODE ANN. § 30-14-104 (2005).
136. Neb. Rev. Stat. § 87-305 (1999).
137. N.H. REV. STAT. ANN. § 358-A:13 (1995).
138. N.M. STAT. §57-12-14 (2005).
139. OHIO. REV. CODE ANN. § 1345.02(c) (LexisNexis 2002).
140. R.I. GEN. LAWS § 6-13.1-3 (2001).
141. S.C. CODE ANN. § 39-5-20 (1985).

Tennessee	To be construed consistently with FTC precedent[142]
Texas	Courts are to be guided by FTC precedent[143]
Utah	Consumer Sales Practices Act: Courts shall construe the Act liberally to promote uniformity with other states enacting it[144]
Vermont	Courts are to be guided by FTC precedent[145]
Virginia	Not stated
Washington	Courts are to be guided by FTC precedent[146]
West Virginia	Courts are to be guided by FTC precedent[147]
Wisconsin	Not stated
Wyoming	Not stated

E. Prohibitions in State Little FTC Acts

Section 5 of the FTC Act states that "[u]nfair methods of competition in or affecting commerce, and unfair or deceptive acts or practices in or affecting commerce, are hereby declared unlawful."[148] By enacting Little FTC Acts, the majority of states and the District of Columbia have adopted the key words from FTC Act Section 5, which are "unfair methods of competition" or "unfair or deceptive acts" or both. Other states have added or substituted language, such as "deceptive acts" without "unfair acts" or "unconscionable acts"; however, this language is arguably in harmony with Section 5 of the FTC Act. Still other states have added an itemized list of proscribed activities, something Congress specifically chose not to include in 1914.[149] Table Two below

142. TENN. CODE ANN. § 47-18-115 (2001).

143. TEX. BUS. & COM. CODE ANN. § 17.46 (c)(1) (Vernon 2002). Courts are also to be guided by § 17.46(b) which gives a nonexclusive set of "false, misleading, or deceptive acts or practices." *Id.* at § 17.46(c)(1).

144. UTAH CODE ANN. § 13-11-2(5) (2001).

145. VT. STAT. ANN. tit. 9, § 2453(b) (1993).

146. WASH. REV. CODE § 19.86.920 (2000).

147. W. VA. CODE ANN. § 46A-6-101 (LexisNexis 1999).

148. 15 U.S.C. § 45(a)(1) (2000).

149. *See* Winerman, *supra* note 28, at 48, 57-58; *see generally* 15 U.S.C. §45 (2000) (reflecting Congress's decision to exclude a detailed list of proscribed acts).

outlines which jurisdictions have included an itemized list of proscribed activities, and which have simply adhered to the traditional language of the FTC Act.

Table 2. Relevant Language Used In Little FTC Acts[150]

STATE	"Unfair Methods of Competition"	"Unfair or Deceptive Acts"	"Deceptive Act"	"Unconscionable Acts"	Detailed List of Specifically Prohibited Acts
Alabama			X	X	X
Alaska	X	X			X
Arizona			X		
Arkansas			X	X	X
California	X	X	X		X
Colorado			X		X
Connecticut	X	X			
Delaware		X[151]	X		X
District of Columbia		X		X	X
Florida	X	X		X	
Georgia		X	X		X
Hawaii	X	X	X		X
Idaho	X	X		X	X
Illinois	X	X	X		X
Indiana			X	X	X
Iowa		X			X
Kansas			X	X	X

150. CCH, *supra* note 91, at ¶ 1200. Copyright 2005 CCH INCORPORATED. All rights reserved. Reprinted with permission from STATE UNFAIR TRADE PRACTICES LAW (Mark D. Bauer ed.). The table has been modified from its original form *See* DEL. CODE ANN. tit. 6, § 2512 (1999); HAW. REV. STAT. § 480-2 (1985); IDAHO CODE ANN. §§ 48-601, 48-603C (2003); ME. REV. STAT. ANN. tit. 5, § 207 (2002); MASS. GEN. LAWS ANN. ch. 93A, § 2 (West 1997); MONT. CODE ANN. § 30-14-101 (2005); S.D. CODIFIED LAWS §§ 37-24-6 (2004); W. VA. CODE ANN. §§ 46A-6-104 (LexisNexis 1999); WIS. STAT. ANN. § 100.18 (West 2004).

151. Delaware's Consumer Fraud Act "unfair or deceptive *merchandising practices*." (emphasis added) DEL. CODE ANN. tit. 6 § 2512 (1999). "Merchandising" is a term of art related to sales, which means the language may not apply to services—most notably the services of learned professionals such as doctors and lawyers.

Kentucky		X		X	
Louisiana	X	X			
Maine	X	X			
Maryland		X			X
Massachusetts	X	X			
Michigan		X	X	X	X
Minnesota			X		X
Mississippi	X	X			X
Missouri			X		
Montana	X	X			
Nebraska	X	X	X	X	X
Nevada			X		X
New	X	X			X
New Jersey			X	X	X
New Mexico		X		X	X
New York			X		
North Carolina	X	X			
North Dakota			X		
Ohio		X	X	X	X
Oklahoma			X		X
Oregon			X	X	X
Pennsylvania	X	X			X
Rhode Island	X	X			X
South Carolina	X	X			
South Dakota			X		X
Tennessee		X			X
Texas			X	X	X
Utah	X		X	X	X
Vermont	X	X			
Virginia			X		X
Washington	X	X			
West Virginia	X	X			
Wisconsin	X		X		X
Wyoming			X		X

Thus, it is clear that Little FTC Acts did not arise independently. The legislative history and the direct input of the FTC indicate that the federal FTC Act was the basis for state consumer protection laws, and, as further proof, the actual language used by many states comes directly from the FTC Act. The question remains whether Little FTC Acts—each with a clear relationship to the FTC Act and, correspondingly, the Sherman Act—could or should be interpreted independently from the FTC Act when the majority of the states have incorporated language requiring that their Little FTC Act be harmonized with federal precedent or the Little FTC Acts of other states.

III. LICENSED PROFESSIONALS UNDER ANTITRUST AND CONSUMER PROTECTION LAWS

The creation of separate and unrelated bodies of precedent for antitrust and consumer protection laws has led to paradoxes that cannot be easily harmonized. One such paradox that illustrates this tension is the way licensed professionals are consistently treated under antitrust law as opposed to how those same professionals are treated under the rubric of consumer protection law.

Since at least 1975 it has been clear that "[t]he nature of an occupation, standing alone, does not provide sanctuary from the Sherman Act"; thus, licensed professionals are subject to the antitrust laws.[152] Beginning, however, with the state of Washington in 1984, courts in several states have held that at least some licensed-professional activities are beyond the scope of the state consumer protection laws.[153]

152. Goldfarb v. Va. State Bar, 421 U.S. 773, 787 (1975).

153. *See* Short v. Demopolis, 691 P.2d 163, 170 (Wash. 1984). A number of state legislatures have exempted specific licensed professionals from consumer protection laws. *See, e.g.*, CAL. CIV. CODE § 1754 (West 1998); FLA STAT. ANN. § 501.212 (West 2002); MD. CODE ANN., COM. LAW § 13-104 (LexisNexis 2005); N.C. GEN STAT. § 75-1.1(b) (2003). Actions taken by a state legislature exempting a specific group is not comparable to a court interpreting a model statute containing language which is identical to the language used in almost every state and in the FTC Act and is, therefore, beyond the scope of this Article.

Table 3. Little FTC Acts And Exemptions Of Licensed Professionals[154]

STATE	CONSUMER PROTECTION STATUTE(S)	EXEMPT PROFESSION(S)	EXEMPTION CREATED BY
Alabama	Deceptive Trade Practices Act[155]	None	N/A
Alaska	Unfair Trade Practices and Consumer Protection Act[156]	None	N/A
Arizona	Arizona Consumer Fraud Act[157]	None	N/A
Arkansas	Deceptive Trade Practices Act[158]	None	N/A
California	Unfair Competition Law[159] and Consumers Legal Remedies Act[160]	Transactions related to real property[161]	Legislature
Colorado	Colorado Consumer Protection Act[162]	None	N/A
Connecticut	Connecticut Unfair Trade Practices Act[163]	Attorneys[164] and Doctors[165]	Court

154. Substantially all states exempt newspapers as well as radio and television stations from responsibility for the content of advertising when the media outlet had no knowledge of the false, misleading or deceptive character of the advertisement; did not prepare the advertisement; and had no financial interest in the sale of the advertised good or service. *See, e.g.*, West Virginia Consumer Protection Act, W. VA. CODE § 46A-5-105 (Lexis 1999).

155. ALA. CODE §§ 8-19-1 to 8-19-15 (LexisNexis 2002).

156. ALASKA STAT. §§ 45.50.471 to 45.50.561 (2004).

157. ARIZ. REV. STAT. ANN. §§ 44-1521 to -1534 (2003).

158. ARK. CODE ANN. §§ 4-88-101 to 4-88-207 (2001).

159. CAL. BUS. & PROF. CODE §§ 17200-17210, 17500-17580 (West 1997).

160. CAL. CIV. CODE §§ 1750-1784 (West 1998).

161. *Id.* at § 1754.

> The provisions of this title shall not apply to any transaction which provides for the construction, sale, or construction and sale of an entire residence or all or part of a structure designed for commercial or industrial occupancy, with or without a parcel of real property or an interest therein, or for the sale of a lot or parcel of real property, including any site preparation incidental to such sale.

Id.

162. COLO. REV. STAT. §§ 6-1-101 to 6-1-115 (2005).

163. CONN. GEN. STAT. ANN. §§ 42-110a to 42-110q (West 2000).

164. Suffield Dev. Assocs. v. Nat'l Loan Investors, 802 A.2d 44, 53 (Conn. 2002) ("[O]nly the entrepreneurial aspects of the practice of law are covered by CUTPA.").

165. Haynes v. Yale-New Haven Hosp., 699 A.2d 964, 974 (Conn. 1997) ("We thus conclude that the touchstone for a legally sufficient CUTPA claim against a health care provider is an allegation that an entrepreneurial or business aspect of the provision of service aside from medical competence is implicated, or aside from medical malpractice . . . ."

Delaware	Consumer Fraud Act and Uniform Deceptive Trade Practices Act[166]	Attorneys[167]	Court
District of Columbia	Consumer Protection Procedures Act[168]	Physicians[169]	Court
Florida	Florida Deceptive and Unfair Trade Practices Act[170]	Licensed Realtors[171]	Legislature
Georgia	Uniform Deceptive Trade Practices Act[172] and Fair Business Practices Act of 1975[173]	None	N/A
Hawaii	Little FTC Act[174] and Uniform Deceptive Trade Practices Act[175]	None	N/A
Idaho	Idaho Consumer Protection Act[176]	None	N/A

166. DEL. CODE ANN. tit. 6, §§ 2511-2527, 2531-2536 (1999).

167. Jamgochian & PKR Enters. v. Prousalis, No. 99 C-10-022, 2000 Del. Super. LEXIS 373, at *11 (Del. Super. Aug. 31, 2000) ("The Constitutional and statutory framework of the regulation of attorneys in this state preclude application of the [Delaware Consumer Fraud] Act, as written, to the conduct of lawyers.").

168. D.C. CODE §§ 28-3901 to -3911 (2001).

169. Dorn v. McTigue, 157 F. Supp. 2d 37, 48 (D.D.C. 2001) ("To separate [consumer protection] claims from standard malpractice claims, [the] court distinguish[es] 'conduct within [sic] the entrepreneurial, commercial, or business aspect of a physician's practice' from the 'actual practice of medicine.'"(quoting Nelson v. Ho, 564 N.W.2d 482, 486 (Mich. Ct. App. 1997))).

170. FLA. STAT. ANN. §§ 501.201 to .213 (West 2002).

171. *Id.* at § 501.212 ("This part does not apply to: . . . (6) An act or practice involving the sale, lease, rental, or appraisal of real estate by a person licensed, certified or registered pursuant to Chapter 475 . . . .").

172. GA. CODE ANN. §§ 10-1-370 to 10-1-375 (2000).

173. *Id.* at §§ 10-1-390 to 10-1-407.

174. HAW. REV. STAT. §§ 480-1 to -24 (1985). The statute is located in the Trade Regulation & Practices section under Monopolies: Restraint of trade in Part I Antitrust Provisions.

175. *Id.* at §§ 481A-1 to -5.

176. IDAHO CODE ANN. §§ 48-601 to -619 (2003).

Illinois	Consumer Fraud and Deceptive Business Practices Act[177] and Uniform Deceptive Trade Practices Act[178]	Licensed Realtors,[179] Attorneys[180] and Doctors[181]	Legislature (Realtors) Court (Attorneys and Doctors)
Indiana	Deceptive Consumer Sales Act[182]	None	N/A
Iowa	Consumer Fraud Act[183]	None	N/A
Kansas	Kansas Consumer Protection Act[184]	None	N/A
Kentucky	Consumer Protection Act[185]	Doctors[186]	Court

177. 815 ILL. COMP. STAT. ANN. 505/1 to 505/12 (West 1999).

178. *Id.* at 510/1 to 510/7.

179. *Id.* at 505/10b(4).

180. Cripe v. Leiter, 703 N.E.2d 100, 107 (Ill. 1998).

> We hold that, where allegations of misconduct arise from a defendant's conduct in his or her capacity as an attorney representing a client, the Consumer Fraud Act does not apply. An attorney's billing of a client for legal services is a part of the attorney's representation of the client and is therefore exempt from the Act.

Id.; *see also* Wafra Leasing Corp. v. Prime Capital Corp., 204 F. Supp. 2d 1120, 1124 (N.D. Ill. 2002) ("Legal services, including the 'actual practice of law' as well as billing practices, are exempt from the Act.") (quoting *Cripe*, 703 N.E.2d at 105)).

181. Gadson v. Newman, 807 F. Supp. 1412, 1417 (C.D. Ill. 1992) ("'The medical and legal professions are afforded immunity from the Consumer Fraud Act primarily, because, unlike other commercial services, medical and legal bodies [sic] are regulated by governmental bodies.'") (quoting Lyne v. Arthur Andersen & Co., 772 F. Supp. 1064, 1068 (N.D. Ill. 1991))).

182. IND. CODE ANN. §§ 24-5-0.5-1 to 24-5-0.5-12 (LexisNexis 1996 & Supp. 2005).

183. IOWA CODE ANN. §§ 714.16 to .16A (West 2003).

184. KAN. STAT. ANN. §§ 50-623 to -643 (1994). The Kansas Supreme Court specifically stated that its Little FTC Act applies to engineering services, although it "ma[d]e no *determination . . . as to application of* the [Little FTC Act] to other professional services." Moore v. Bird Eng'g Co., 41 P.3d 755, 761-62 (Kan. 2002).

185. KY. REV. STAT. ANN. §§ 367.110 to .990 (LexisNexis 2002).

186. Simmons v. Stephenson, 84 S.W.3d 926, 927-28 (Ky. Ct. App. 2002).

> We agree with the trial court that, while the [plaintiffs'] complaint may state a cause of action for medical malpractice, the [plaintiffs] may not pursue their claims under the Consumer Protection Act . . . . In our opinion eye surgery does not constitute a '*trade*' or '*commerce*' within the meaning of the statute.

Id. at 927. According to the court in *Simmons*, in order for their Consumer Protection Act to apply, the claim must relate to the entrepreneurial, commercial, or business aspect of the defendant's practice of medicine. *Id.* at 928.

Louisiana	Unfair Trade Practices and Consumer Protection Law[187]	Attorneys[188]	Court
Maine	Maine Unfair Trade Practices Act[189] and Uniform Deceptive Trade Practices Act[190]	None	N/A
Maryland	Consumer Protection Act[191]	Attorneys, CPAs, Vets, Chiropractors, Physical Therapists, Optometrists, Land Surveyors, Architects, Clergymen, Engineers, Insurance Providers, Podiatrists, Medical or Dental Practitioners, and Real Estate Brokers or Salespersons[192]	Legislature
Massachusetts	Consumer Protection Act[193]	None	N/A
Michigan	Michigan Consumer Protection Act[194]	Doctors[195] and Attorneys[196]	Court

187. LA. REV. STAT. ANN. §§ 51:1401 to :1420 (2003).

188. Jackson v. Adcock, No. 03-3369, 2004 U.S. Dist. LEXIS 16888, at *19 (E.D. La., Aug. 23, 2004) ("LUTPA *does not regulate the practice of law.*").

189. ME. REV. STAT. ANN. tit. 5, §§ 205A to 214 (2002).

190. *Id.* at tit. 10, §§ 1211-1216 (1997).

191. MD. CODE ANN. COM. LAW §§ 13-101 to -501 (2005).

192. The Maryland Consumer Protection Act does not apply to the professional services of a certified public accountant, architect, clergyman, professional engineer, lawyer, veterinarian, insurance company authorized to do business in the State, insurance producer licensed by the State, Christian Science practitioner, land surveyor, property line surveyor, chiropractor, optometrist, physical therapist, podiatrist, real estate broker, associate real estate broker, or real estate salesperson, or medical or dental practitioner.

Id. at § 13-104.

193. MASS. GEN. LAWS ANN. ch. 93A, §§1-11 (West 1997).

194. MICH. COMP. LAWS ANN. §§ 445.901 to .922 (West 2002).

195. Nelson v. Ho, 564 N.W.2d 482, 486 (Mich. Ct. App. 1997) ("Only when physicians are engaging in the entrepreneurial, commercial, or business aspect of the practice of medicine are they engaged in 'trade or commerce' within the purview of the MCPA.").

196. Comer Family Equity v. Thomas & Jensen P.C., No. 186676, 1997 Mich. App. LEXIS 2949, at *9 (Mich. Ct. App. Apr. 25, 1997) ("We adopt this Court's holding in *Nelson* as our own, and apply it to the practice of law . . . ." (citing *Nelson*, 564 N.W.2d at 486)).

Minnesota	Uniform Deceptive Trade Practices Act;197 Consumer Fraud Act;[198] and Duties of Attorney General (Consumer Affairs)[199]	None	N/A
Mississippi	Consumer Protection Act[200]	None	N/A
Missouri	Merchandising Practices Act[201]	None	N/A
Montana	Montana Unfair Trade Practices and Consumer Protection Act of 1973[202]	None	N/A
Nebraska	Consumer Protection Act[203] and Uniform Deceptive Trade Practices Act[204]	None	N/A
Nevada	Deceptive Trade Practices Act[205]	None	N/A
New Hampshire	Consumer Protection Act[206]	Attorneys[207] and Doctors[208]	Court

197. MINN. STAT. ANN. § 8.31 (West 2005); MINN. STAT. ANN. §§ 325D.43 to .48 (West 2004).

198. MINN. STAT. ANN. §§ 325F.67 to .71 (West 2004).

199. MINN. STAT. ANN. §§ 8.31 to .32 (West 2005).

200. MISS. CODE ANN. §§ 75-24-1 to 75-24-27 (2000).

201. MO. ANN. STAT. §§ 407.010 to .307 (West 2001).

202. MONT. CODE ANN. §§ 30-14-101 to 30-14-143 (2005).

203. NEB. REV. STAT. §§ 59-1601 to -1623 (2004).

204. *Id.* at §§ 87-301 to -306 (1999).

205. NEV. REV. STAT. §§ 598.0903 to .0999 (1997).

206. N.H. REV. STAT. ANN. §§ 358-A:1 to :13 (1995).

207. Rousseau v. Eshleman, 519 A.2d 243, 245 (N.H. 1986) ("[A]bsent careful legislative consideration and a clearly expressed legislative intent" attorneys are exempt from the New Hampshire Consumer Protection Act.); *see also* Averill v. Cox, 761 A.2d 1083, 1089 (N.H. 2000) ("[D]isciplinary measures exercised [by the New Hampshire Supreme Court against attorneys] protect the public as effectively from deceptive or unfair actions in the marketplace as would double or treble damages under [the New Hampshire Consumer Protection] Act.").

208. *Rousseau*, 519 A.2d at 245 (dicta) ("Presumably, physicians would be considered exempt from the act . . . .").

New Jersey	Consumer Fraud Act[209]	Hospitals,[210] Learned Professionals,[211] and Attorneys[212]	Court
New Mexico	Unfair Practices Act[213]	None	N/A
New York	Deceptive Acts and Practices Act[214] and Enforcement by Attorney General[215]	None	N/A
North Carolina	Unfair Trade Practices Act[216]	Members of learned professions rendering professional services[217]	Legislature
North Dakota	Consumer Fraud Act[218]	None	N/A
Ohio	Consumer Sales Practices Act[219] and Deceptive Trade Practices Act[220]	Accountants, Attorneys, Physicians, Dentists, Veterinarians,[221] and Chiropractors[222]	Legislature

209. N.J. STAT. ANN. §§ 56:8-1 to -92 (West 2001).

210. Hampton Hosp. v. Bresan, 672 A.2d 725, 731 (N.J. Super. Ct. App. Div. 1996) ("[The court] can find no purpose to a requirement that hospital services be within the purview of the Consumer Fraud Act . . . .").

211. Macedo v. Dello Russo, 840 A.2d 238, 242 (N.J. 2004). "[F]orty years after the CFA was enacted, our jurisprudence continues to identify learned professionals as beyond the reach of the Act so long as they are operating in their professional capacities. The Legislature is presumed to be aware of that judicial view." *Id.*

212. Vort v. Hollander, 607 A.2d 1339, 1342 (N.J. Super. Ct. App. Div. 1992). ("[A]ttorney's services do not fall within the intendment of the Consumer Fraud Act.").

213. N.M. STAT. §§ 57-12-1 to 57-12-24 (2005).

214. N.Y. GEN. BUS. LAW §§ 349-350 (McKinney 2004).

215. N.Y. EXEC. LAW § 63(12) (McKinney 2002).

216. N.C. GEN STAT. §§ 75-1.1 to -38 (2003).

217. *Id.* at § 75-1.1(b) ("[This section] does not include professional services rendered by a member of a learned profession.").

218. N.D. CENT. CODE §§ 51-15-01 to 51-15-11 (1999).

219. OHIO REV. CODE ANN. §§ 1345.01 to .13 (LexisNexis 2002).

220. OHIO REV. CODE ANN. §§ 4165.01 to .04 (LexisNexi 2001).

221. "'Consumer transaction' does not include . . . transactions between certified public accountants or public accountants and their clients; transactions between attorneys, physicians, or dentists and their clients or patients; and transactions between veterinarians and their patients that pertain to medical treatment but not ancillary services." *Id.* at § 1345.01(A).

222. Chiropractic Clinic of Solon v. Kutsko, 636 N.E.2d 422, 423-24 (Ohio Ct. App. 1994) ("[W]e conclude that licensed doctors of chiropractic are 'physicians' as the term is used in [the Consumer Sales Practices Act section] 1345.01(A).").

Oklahoma	Consumer Protection Act[223] and Oklahoma Deceptive Trade Practices Act[224]	Those whose professions are regulated under state or federal law [225]	Court
Oregon	Unlawful Trade Practices Act[226]	None	N/A
Pennsylvania	Unfair Trade Practices and Consumer Protection Law[227]	Physicians[228] and Lawyers[229]	Court
Rhode Island	Unfair Trade Practice and Consumer Protection Act[230]	None	N/A
South Carolina	South Carolina Unfair Trade Practices Act[231]	None	N/A
South Dakota	Deceptive Trade Practices and Consumer Protection Law[232]	None	N/A
Tennessee	Tennessee Consumer Protection Act of 1977[233]	Doctors[234]	Court

223. OKLA. STAT. ANN. tit. 15, §§ 751-763 (West 1993)

224. OKLA. STAT. ANN. tit. 78, §§ 51-55 (West 2002).

225. OKLA. STAT. ANN. tit. 15, § 754(2) ("Nothing in this act shall apply to . . . [a]ctions or transactions regulated under . . . any . . . regulatory body or officer acting under statutory authority of this state of the United States . . . ."); *see, e.g.*, Estate of Hicks v. Urban East, Inc., 92 P.3d 88, 94-95 (Okla. 2004) ("[T]he provision of care and medical services by nursing homes is regulated under laws administered by the Oklahoma Department of Health . . . , and as such falls within the exemption to the Consumer Protection Act . . . § 754(2).").

226. OR. REV. STAT. §§ 646.605 to .656 (1999).

227. 73 PA. STAT. ANN. §§ 201-1 to -9.3 (West 1993 & Supp. 1997).

228. Gatten v. Merzi, 579 A.2d 974, 976 (Pa. Super. Ct. 1990) ("It . . . is clear that the legislature did not intend the [Unfair Trade Practices and Consumer Protection Law] to apply to physicians rendering medical services.").

229. *See* SHELDON & CARTER *supra* note 101, § 2.3.9 at 67 ("Federal Courts ruling on UDAP issues in Pennsylvania have noted that coverage of attorneys may be inconsistent with the state supreme court's exclusive constitutional authority to regulate the conduct of attorneys, but no court has yet decided the issue.").

230. R.I. Gen. Laws §§ 6-13.1-1 to 6-13.1-27 (2001).

231. S.C. CODE ANN. §§ 39-5-10 to 39-5-170 (1985).

232. S.D. CODIFIED LAWS §§ 37-24-1 to 37-24-35 (2004).

233. TENN. CODE ANN. §§ 47-18-101 to 47-18-125 (2001).

234. Constant v. Wyeth, Inc., 352 F. Supp. 2d 847, 853-54 (M.D. Tenn. 2003) ("[M]edical malpractice claims may not be recast as consumer protection act claims.").

Texas	Deceptive Trade Practices-Consumer Protection Act[235]	Professional service providers[236]	Legislature
Utah	Unfair Practices Act[237] and Utah Consumer Sales Practices Act[238]	None	N/A
Vermont	Consumer Fraud Act[239]	Attorneys[240]	Court
Virginia	Virginia Consumer Protection Act of 1977[241]	Health Care and abortion Providers[242] and Real Estate Licensees[243]	Court (Health Care Providers) Legislature (Real Estate Licensees)
Washington	Consumer Protection Act[244]	Doctors[245] and Attorneys [246]	Court
West Virginia	Consumer Credit and Protection Act[247]	None	N/A

235. TEX. BUS. & COM. CODE ANN. §§ 17.41 to .63 (Vernon 2002).

236. *Id.* at § 17.49c. "Nothing in this subchapter shall apply to a claim for damages based on the rendering of a professional service, the essence of which is the providing of advice, judgment, opinion, or similar professional skill." *Id.*

237. UTAH CODE ANN. §§ 13-2-1 to 13-2-9, 13-5-1 to 13-5-18 (2005).

238. *Id.* at §§ 13-11-1 to13-11-23.

239. VT. STAT. ANN. tit. 9, §§ 2451-2480g (1993).

240. Kessler v. Loftus, 994 F. Supp. 240, 242-43 (D. Vt. 1997) ("[R]epresentations made by an attorney to a client are exempt from the CFA if they fall within the definition of the actual practice of law. . . . [defined as] those matters 'requiring the professional judgment of an attorney based upon his or her legal knowledge and skill.'" (second alteration in original) (quoting Roussea v. Eshleman, 579 A.2d 243, 250 (1986))).

241. VA. CODE ANN. §§ 59.1-196 to -207 (2001).

242. Ott v. Baker, 53 Va. Cir. 113, 115 (Va. Cir. Ct. 2000) ("It is clear that health care and abortions fall within the exclusion set forth in [the Virginia Consumer Protection Act], § 59.1-199.").

243. VA. CODE ANN. §59.1-199(F) (2001) (stating that real estate licensees licensed under Chapter 21 of Title 54.1 are exempt).

244. WASH. REV. CODE §§ 19.86.010 to 19.86.920 (2000).

245. Quimby v. Fine, 724 P.2d 403, 406 (Wa. Ct. App. 1986). "We see no basis to distinguish the legal practice from the medical practice. Therefore the [plaintiff's medical] negligence claim paralleling legal negligence . . . is not under the Consumer Protection Act because it relates to the actual competence of the medical practitioner." *Id.* (citing Short v. Demopolous, 691 P.2d 163 (1984)).

246. Eriks v. Denver, 824 P.2d 1207, 1214 (Wa. 1992) ("The provision of legal services does not generally fall within the definition of 'trade or commerce,' except as those services relate to the 'entrepreneurial aspects' of the practice of law.").

247. W. VA. CODE ANN. §§ 46A-6-101 to 46A-6-110 (LexisNexis 1999).

Wisconsin	Fraudulent Representations[248] and Unfair Trade Practices Act[249]	Licensed Real Estate Brokers and Salespersons[250]	Legislature
Wyoming	Wyoming Consumer Protection Act[251]	None	N/A

A. Licensed Professionals Under the Sherman Act

Section 1 of the Sherman Act declares illegal "[e]very contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States, or with foreign nations."[252] Unfortunately, antitrust case law left unresolved for decades the question of whether licensed professions constitute a trade.[253] In fact, as recently as 1950,

---

248. WIS. STAT. ANN. § 100.18 (West 2004).

249. *Id.* at §§ 100.20 to .264.

250. *Id.* at § 100.18(12)(b) (exempting licensed real estate brokers and salespersons without knowledge of falsity).

251. WYO. STAT. ANN. §§ 40-12-101 to 40-12-114 (2005).

252. 15 U.S.C. § 1 (2000). The antitrust laws are derived from the commerce clause of the United States Constitution. WILLIAM LETWIN, LAW AND ECONOMIC POLICY IN AMERICA 89 (1965). While this Article focuses on the word "trade," the definition of "commerce" may also be of interest. There are two approaches used by courts to determine whether a challenged practice has the necessary effect on interstate commerce. McLain v. Real Estate Bd. of New Orleans, Inc., 444 U.S. 232, 241-42 (1980). The two tests are the "flow of commerce" test and the "effecting commerce" test. *See id.* (referring to the tests as "in commerce" and "effect on commerce" theories). In the "flow of commerce" test, one sale of a product over state lines meets the Sherman Act interstate commerce requirement. *See* Apex Hosiery Co. v. Leader, 310 U.S. 469, 495, 517 (1940) (noting that it is the nature of the restraint and its effect on interstate commerce which are key, not the amount of the commerce). Even if the allegedly unlawful conduct is largely intrastate in nature, it may still meet the Sherman Act's trade or commerce requirement under the "effecting commerce" test. Hosp. Bldg. Co. V. Trustees of Rex Hosp., 425 U.S. 738, 743 (1976). Under the "effecting commerce" test, a plaintiff must show that the defendant's general business activity has an effect on interstate commerce, not that the restraint itself affects interstate commerce. *Id.* The Supreme Court has held that Congress intended the Sherman Act to be exercised to the fullest constitutional extent under the commerce clause. HOVENKAMP, *supra* note 13, at 747.

253. *See, e.g.*, FTC v. Raladam Co., 283 U.S. 643, 653 (1931) ("[M]edical practitioners .... follow a profession and not a trade...."); Fed. Baseball Club of Balt. v. Nat'l League of Prof'l Baseball Clubs, 259 U.S. 200, 209 (1922) ("[A] firm of lawyers sending out a member to argue a case ... does not engage in ... commerce because the lawyer ... goes to another State."); *see also* United States v. Nat'l Ass'n of Real Estate Bds., 339 U.S. 485, 490-91 (1950) ("'Wherever any occupation ... is carried on for the purpose of profit, gain, or livelihood, not in the liberal arts or in the learned professions, it is constantly called trade." (quoting The Nymph, 18 F. Cas. 506 (D. Me.) (No. 10,388))); Atlantic Cleaners & Dyers, Inc. v. United States, 286 U.S. 427, 436 (1932) ("Wherever any occupation, employment, or business is

the Supreme Court expressly refused to rule on the question.[254]

The Supreme Court resolved the question in *Goldfarb v. Virginia State Bar*.[255] In *Goldfarb*, a home buyer sued the Virginia State Bar alleging that a minimum fee schedule for attorneys examining titles violated Section 1 of the Sherman Act.[256] The Virginia Bar argued that Congress never intended the Sherman Act to regulate the "learned professions" because "competition is inconsistent with the practice of a profession because enhancing profit is not the goal of professional activities; the goal is to provide services necessary to the community."[257] In its argument, the Virginia Bar cited to dicta in *Raladam* that stated "medical practitioners . . . follow a profession and not a trade."[258]

According to the Supreme Court, the "classic" argument to exempt licensed professionals from the Sherman Act was that "because enhancing profit is not the goal of professional activities . . . . [professions are distinguished] from trades, businesses, and other occupations."[259]

Noting that the Virginia Bar's argument lost some of its force when used to support price-fixing, the court ruled that "Congress intended to strike as broadly as it could under § 1 [sic] of the Sherman Act, and to read into it so wide an exception as that urged on us would be at odds with that purpose."[260] The court further emphasized that neither the "nature of an occupation" nor the "public-service aspect of professional practice" would exempt conduct from the antitrust laws.[261]

The *Goldfarb* Court resolved any lingering doubt that the Sherman Act applies to the activities of licensed professionals, and that any distinction between the words "trade," or "trade or commerce," and "profession" has no relevance under the Sherman Act.[262] Since then, courts applying *Goldfarb*

---

carried on for the purpose of profit, or gain, or a livelihood, not in the liberal arts or in the learned professions, it is constantly called a *trade*."). The Supreme Court noted, however, that these statements were passing references in cases concerned with other issues. Goldfarb v. Va. State Bar, 421 U.S. 773, 786 n.15 (1975).

254. *National Ass'n of Real Estate Bds.*, 339 U.S. at 491-92 ("We do not intimate an opinion on the correctness of the application of the term ["trade"] to the professions.").

255. 421 U.S. 773 (1975).

256. *Id.* at 775-78. In fact, charging less was presumed by the association to be a form of ethical misconduct. *Id.* at 777-78.

257. *Id.* at 786.

258. *Id.* at 786 n.15 (citing *Raladam*, 283 U.S. at 653).

259. *Goldfarb*, 421 U.S. at 786-87 (footnote omitted).

260. *Id.* at 787. The court also noted that "[t]he reason for adopting the fee schedule does not appear to have been wholly altruistic. The first sentence in respondent State Bar's 1962 Minimum Fee Schedule Report states: 'The lawyers have slowly, but surely, been committing economic suicide as a profession.'" *Id.* at 786 n.16 (citation omitted).

261. *Id.* at 787. The Supreme Court has since clarified that there is no learned professions exemption under the Sherman Act. Nat'l Soc'y of Prof'l Eng'rs v. United States, 435 U.S. 679, 696 (1978) ("The cautionary footnote in *Goldfarb* . . . cannot be read as fashioning a broad exemption under the Rule of Reason for learned professions." (citing *Goldfarb*, 421 U.S. at 788 n.17)).

262. *Goldfarb*, 421 U.S. at 787-88.

have held that the Sherman Act applies to substantially all professional activities, even philanthropic professional activities.[263]

B. Licensed Professionals Under the FTC Act

Case law and other precedent clearly indicate that licensed professionals are within the scope of the FTC Act.[264] The FTC Act expressly exempts certain classes of persons from its coverage, including certain financial institutions and air carriers, however, the list does not include any of the learned professions.[265] When the Supreme Court, in 1975, ruled that licensed professionals were within the scope of the Sherman Act, it was clear that licensed professionals were also within the scope of the FTC Act. Although states, as grantors of most professional licenses, appear to engage in more litigation involving professional activities, the FTC has been involved as well.[266]

In 1979, the FTC brought suit against the American Medical Association ("AMA") for violating Section 5 of the FTC Act by creating and enforcing advertising, contract, and solicitation restraints on physicians and between physicians and non-physicians.[267] The AMA argued that, as a nonprofit corporation with the objective of safeguarding the medical profession, its activities were not subject to the jurisdiction of the FTC Act.[268] Citing *Goldfarb*,[269] the United States Court of Appeals for the Second Circuit

---

263. *See, e.g.*, NCAA v. Bd. of Regents of Univ. of Okla., 468 U.S. 85, 100 n.22 (1984) (noting that the Sherman Act applies to non-profit entities); Arizona v. Maricopa County Med. Soc'y, 457 U.S. 332, 339, 357 (1982) (holding that a nonprofit organization of doctors violated the antitrust laws by setting a maximum fee schedule); Am. Soc'y of Mech. Eng'rs, Inc. v. Hydrolevel Corp., 456 U.S. 556, 576 (1982) ("[N]onprofit organizations can be held liable under the antitrust laws."); Nat'l Soc'y of Prof'l Eng'rs v. United States, 435 U.S. 679, 681 (1978) (holding that a nonprofit association's rule against competitive bidding violated antitrust laws); United States v. Brown Univ., 5 F.3d 658, 668 (3d Cir. 1993) (concluding that the act of setting tuition and fees is trade or commerce under the Sherman Act) (citing *Goldfarb*, 421 U.S. at 788 n.17)); Ass'n for Intercollegiate Athletics for Women v. NCAA, 735 F.2d 577, 582-83 (D.C. Cir. 1984) (rejecting the "NCAA'[s] argu[ment] that its nonprofit status and affiliation with higher education warrants special treatment under the antitrust laws").

264. *See* Heslin v. Conn. Law Clinic of Trantolo & Trantolo, 461 A.2d 938, 942-43 (Conn. 1983) (utilizing case law and Federal Trade Commission interpretations of the FTC Act to conclude that the FTC Act applies to attorneys).

265. 15 U.S.C. § 45(a)(2) (2000).

266. *See, e.g.*, Cal. Dental Ass'n v. FTC, 224 F.3d 942, 958 (9th Cir. 2000) (issuing remand for FTC to apply Sherman Act rule of reason analysis to establish a violation of Section 5 of the FTC Act); FTC v. Ind. Fed'n of Dentists, 476 U.S. 447, 451-52, 455 (1986) (using a Sherman Act violation to establish a violation of Section 5 of the FTC Act).

267. Am. Med. Ass'n v. FTC, 638 F.2d 443, 445-46 (2d Cir. 1980), *aff'd* 455 U.S. 676, 676 (1982) (per curiam).

268. *Id.* at 447-48.

269. Goldfarb v. Va. State Bar, 421 U.S. 773, 778 (1975).

disagreed and held that the AMA's activities were within the scope of the FTC Act.[270] The Supreme Court affirmed.[271]

In addition, the FTC has repeatedly stated, during litigation and in advisory opinions, that it has the power to regulate state-licensed and regulated professionals under Section 5 of the FTC Act. In *In re Wilson Chemical Co.*,[272] the FTC considered whether collection efforts by an attorney were outside the scope of the FTC Act because "the collection letter of an attorney is not commerce within the Federal Trade Commission Act."[273] The FTC held that the attorney was as equally liable for a violation of Section 5 of the FTC Act as the collection company he employed to send out the letters.[274] The FTC has also unequivocally stated its official position that state-regulated professions are not, and should not be, exempted from the FTC Act.[275]

Congress has considered and rejected proposals to strip the FTC of its power to regulate professionals. In 1982, Congress considered an amendment to the FTC Act which would have provided a specific exemption for "practitioners of a profession whose members are licensed and regulated by a State as a condition of independent practice within the State."[276] Congress voted down the proposal[277] and it has not resurfaced.

C. Licensed Professionals Under State Little FTC Acts

Whether the learned professions, such as doctors, lawyers, accountants, architects, and engineers[278] should be included under Little FTC Acts has been debated for years; the question parallels the debate over whether these professionals were within the scope of the Sherman Act. Statements made in advertising are most likely to create tension between a licensed professional and a Little FTC Act. Until recently, the professions were generally banned by state law or other codes of conduct from advertising.[279] Since the Supreme

---

270. *Am. Med. Ass'n*, 638 F.2d at 448. The Second Circuit also cited FTC v. Nat'l Comm'n on Egg Nutrition, 517 F.2d 485, 488 (7th Cir. 1975), *cert. denied*, 426 U.S. 919 (1976) ("Congress did not intend to provide a blanket exclusion of all non-profit corporations . . . .").

271. *Am. Med. Ass'n*, 455 U.S. at 676. The Court also denied a petition for rehearing. Am. Med. Ass'n v. FTC, 456 U.S. 966, 966 (1982).

272. 64 F.T.C. 168 (1964).

273. *Id.* at 186.

274. *See id.* at 187.

275. Reauthorization of the Federal Trade Commission, 1982 Hearing on S. 2499, Before the S. Comm. on Commerce, Sci., and Transp., 97th CONG. REC., 31,387 (1982) (letter from Comm'r James C. Miller III, Chairman).

276. S. 2499, 97th Cong. § 3(c) (2d Sess. 1982).

277. S. 2499, 97th Cong. §3(c) (2d Sess. 1982), 128 CONG. REC. 31,387-89 (1982); *see also* SEN. REP. NO. 98-215, at 10-11 (1983).

278. Some states also include real estate professionals. *See, e.g.*, Florida Unfair and Deceptive Trade Practices Act, FLA. STAT. ANN. § 501.212 (West 2002) (specifying an exemption for real estate "sale, lease, rental or appraisal by a person licensed" under state law).

279. PRIDGEN, *supra* note 92, § 4:34.

Court's decision in *Goldfarb*,[280] holding that the professions could be liable for anticompetitive conduct under the Sherman Act, many state courts have ruled that the professions are also within the scope of consumer protection laws.[281] The state courts which have found some or all professionals to be outside the scope of their Little FTC Act are described below and categorized by the court's rationale; some courts have suggested more than one reason to exempt professionals.

The decisions described below unfortunately lack some depth and detail, but the court decisions themselves are often conclusory with regard to interpretation or explanation of the relevant Little FTC Act.

i. *Trade or Commerce Exemptions*

Similar to the FTC and Sherman Acts, the majority of states require that allegedly unlawful conduct under consumer protection laws must be in "trade or commerce."[282] Substantially all of the remaining states require that the offending conduct arise from "trade."[283] The Washington Supreme Court held that an attorney's conduct in the practice of law may not be considered "trade or commerce."[284] A billing dispute arose between a law firm and client concerning both the size of the bill and whether the client had agreed that an associate and a younger partner would work on the case, rather than the controlling partner.[285] The client alleged a violation of Washington's Little FTC Act.[286]

The Washington Supreme Court made a comprehensive analysis, with one critical flaw: the court assumed that, prior to *Goldfarb*,[287] the antitrust laws exempted licensed professionals.

> Plaintiffs . . . assert that [since the Washington Little FTC Act] was adopted virtually verbatim from federal antitrust laws, the Legislature is presumed to have intended 'trade or commerce' to be interpreted in accordance with then-existing construction of that term under federal law. Since before 1961 [when the Washington Little FTC Act was enacted], federal law did not consider the learned professions to be part of trade or commerce; ergo, the

---

280. Goldfarb v. Va. State Bar, 421 U.S. 773 (1975).

281. PRIDGEN, *supra* note 92, § 4:34.

282. *See, e.g.*, 73 PA. STAT. ANN. § 201-3 (West 1993). The FTC Act only references "commerce," but the definition of commerce in the Act is substantively identical to the definition of "trade and commerce" in the Sherman Act. *Compare* FTC Act, 15 U.S.C. § 44 (defining commerce without using the word trade), *with* Sherman Act, 15 U.S.C. § 12(a) (defining commerce using the word trade).

283. CCH *supra* note 91, at ¶ 1410.

284. Short v. Demopolis, 691 P.2d 163, 170 (Wash. 1984).

285. *Id.* at 164.

286. *Id.* at 165.

287. Goldfarb v. Va. State Bar, 421 U.S. 773 (1975).

> practice of law cannot constitute trade or commerce under the [Washington Little FTC Act].[288]

The Washington Supreme Court appears to have erred. In *Goldfarb*, the Supreme Court held that the issue of whether licensed professionals were exempt under the antitrust laws was one of first impression and any previous discussion was one of "passing references in cases concerned with other issues . . . until the present case it is clear that we have not attempted to decide whether the practice of a learned profession falls within [Section] 1 of the Sherman Act."[289] Although the Washington Supreme Court did hold that "certain entrepreneurial aspects of the practice of law may fall within the 'trade or commerce' definition" of the Little FTC Act, it refused to recognize that all attorney conduct was trade or commerce.[290]

Other courts have indulged in more creative analyses. In trying to distinguish antitrust cases concerning the learned professions, the Illinois Court of Appeals suggested that such cases "dealt only with the commercial aspects of the legal profession through activities which would have a direct effect on the consuming public and not with the practice of law itself."[291] Although the court failed to describe the type of activity which would involve the practice of law but not have a direct effect on the consuming public, the court held that "trade or commerce" did not include the actual practice of law.[292]

---

288. *Short*, 691 P.2d at 168.

289. *Goldfarb*, 421 U.S. at 786 n.16 (citing United States v. Nat'l Ass'n of Real Estate Bds., 339 U.S. 485, 492 (1950)); Am. Med. Ass'n v. United States, 317 U.S. 519, 528 (1943).

290. *See Short*, 691 P.2d at 168 ("[C]laims . . . concern[ing] the actual practice of law. . . . amount[ing] to allegations of negligence or malpractice . . . are exempt from the [Little FTC Act]."). Justice Pearson's concurring opinion relied on the discredited case of *Marjorie Webster Junior College, Inc. v. Middle States Association of Colleges & Secondary Schools, Inc.*, 432 F.2d 650 (D.C. Cir. 1970) *cert. denied*, 400 U.S. 965 (1970). *Short*, 691 P.2d at 171 (Pearson, J. concurring). *Marjorie Webster* held that non-profit organizations were exempt from the antitrust laws; while never directly reversed, *Marjorie Webster* is generally not considered to be good law. *See Marjorie Webster*, 432 F.2d at 654; *see, e.g.*, NCAA v. Bd. of Regents of the Univ. of Okla., 468 U.S. 85, 100 n.22 (1984) (noting that the Sherman Act applies to non-profit entities); Arizona v. Maricopa County Med. Soc'y, 457 U.S. 332, 339, 357 (1982) (holding that a nonprofit organization of doctors violates the antitrust laws by setting a maximum fee schedule); Am. Soc'y of Mech. Eng'rs, Inc. v. Hydrolevel Corp., 456 U.S. 556, 576 (1982) ("[N]onprofit organizations can be held liable under the antitrust laws."); Nat'l Soc'y of Prof'l Eng'rs v. United States, 435 U.S. 679, 696-97 (1978) (determining that a nonprofit association's rule against competitive bidding violated antitrust laws).

291. Frahm v. Urkovich, 447 N.E.2d 1007, 1010 (Ill. App. Ct. 1983).

292. *Id.* at 1011; *see also* Lurz v. Panek, 527 N.E.2d 663, 665, 670 (Ill. App. Ct. 1988) (refusing to apply Illinois Little FTC Act even though alleged wrongful conduct, "failure to disburse promptly funds obtained . . . in an underlying . . . action," was not integral to the "actual practice of law.").

ii. *Non-Entrepreneurial Activities Exemptions*

Several state courts have created a subjective test to determine the applicability of a Little FTC Act. If a licensed professional is engaged in an entrepreneurial activity, then the conduct falls within the ambit of the Little FTC Act. If the activity involves the learned profession itself, then the Little FTC Act does not apply. Three examples of state courts following the "actual practice" of the profession exception are given below.

In Vermont, a law firm represented to a divorce client that a mortgage on her former spouse's land provided "adequate security," and that he had represented her interest competently.[293] She claimed that she did not receive either.[294] Although the court noted that it was required to construe Vermont's Little FTC Act in accordance with FTC precedent, and that attorneys received no blanket exemption from the law, the court held that representations of "adequate security" and "competent representation" were legal opinions and not entrepreneurial in nature.[295]

In Connecticut, a debtor alleged that a law firm fraudulently and deceptively tried to collect a debt.[296] In a separate count, the Supreme Court of Connecticut agreed that the law firm abused process.[297] The court, however, denied relief under the Little FTC Act.[298] The court concluded that, although seeking to recover an amount in excess of what was owed may have been professional misconduct, the alleged activity did not pertain to entrepreneurial aspects of the practice of law; only activities pertaining to entrepreneurial aspects are actionable under the Little FTC Act.[299]

In Tennessee, a doctor prescribed the drug Fen-Phen to a patient; the drug was later withdrawn from "the market because of concerns about serious health effects."[300] The court succinctly held that doctors are immune from Tennessee's Little FTC Act when the allegations concern the actual provision of medical services.[301]

---

293. Kessler v. Loftus, 994 F. Supp. 240, 241 (D. Vt. 1997).

294. *Id.*

295. *See id.* at 242, 244.

296. Suffield Dev. Assocs. v. Nat'l Loan Investors, L.P., 802 A.2d 44, 47 (Conn. 2002).

297. *Id.* at 49.

298. *Id.* at 53 ("[W]e conclude that these allegations do not support a [Little FTC Act] claim because obtaining an execution on a judgment relates to the representation by the defendant attorney and law firm of their client and not to the entrepreneurial aspects of practicing law.").

299. *Id.*

300. Constant v. Wyeth, 352 F. Supp. 2d 847, 849 (M.D. Tenn. 2003).

301. *See id.* at 853-54 & n.10 ("'[O]nly allegations of unfair, unconsionable, or deceptive methods, acts, or practices in the conduct of the entrepreneurial, commercial, or business aspect of a physician's practice may be brought under the [Little FTC Act].'" (quoting Nelson v. Ho, 564 N.W.2d 482, 486 (Mich. Ct. App. 1997))).

iii. *Regulated Professions Exemptions*

Another reason that some state courts have chosen to exempt licensed professionals stems from the license itself. Some state courts yield to the regulatory scheme already in place for licensed professionals,[302] and at least one state court found that a vaguely worded portion of its model act exempts attorneys.[303]

In Illinois, a psychiatrist accused a hospital and another psychiatrist of deceptively creating financial incentives for patient admissions.[304] While the court acknowledged that state-regulated professionals were not exempt from the FTC Act itself and that the Illinois Little FTC Act called upon courts to consult FTC precedent, the court held that "'[t]he medical and legal professions are afforded immunity from the [Illinois Little FTC Act] primarily, because, unlike other commercial services, medical and legal bodies [sic] are regulated by governmental bodies.'"[305]

A New Jersey plaintiff alleged that a hospital inflated its medical bills by unnecessarily extending the plaintiff's stay.[306] Holding hospitals to be beyond the scope of the New Jersey Little FTC Act, the court noted that hospitals were already strongly regulated by the state's Department of Health.[307] The court did not note, however, whether this separate regulatory scheme included a right to private action or to multiple damages.

The New Hampshire Supreme Court wrote "[o]ur holding is supported by the legislature's failure to expressly include the practice of law within the Act's scope since our [first decision on this point] was issued almost fifteen years ago."[308] The court had previously decided that attorneys and other professionals were exempt from New Hampshire's Little FTC Act because of the vague wording exempting trade or commerce which was subject to a

---

302. *See, e.g.*, Cripe v. Leiter, 703 N.E.2d 100, 107 (Ill. 1998) (An attorney's billing of a client for legal services is a part of the attorney's representation of the client[, is covered by the regulatory scheme for attorney's,] and is therefore exempt from the [Little FTC] Act.").

303. Averill v. Cox, 761 A.2d 1083, 1087-88 (N.H. 2000) ("We conclude that our regulation of the practice of law is comprehensive and protects consumers from the same fraud and unfair practices as [the Little FTC Act].").

304. Gadson v. Newman, 807 F. Supp. 1412, 1414 (C.D. Ill. 1992).

305. *Id.* at 1417-19 (quoting Lyne v. Arthur Andersen & Co., 772 F. Supp. 1064, 1068 (N.D. Ill. 1991)). The court, however, allowed the plaintiff the opportunity to prove his consumer fraud case "because contracts between psychiatric clinics and hospitals are [not] subject to extensive regulation" and "the instant case falls outside the expertise of the professional regulations of the medical community." *Id.*

306. *See* Hampton Hosp. v. Bresan, 672 A.2d 725, 726 (N.J. Super. Ct. App. Div. 1996).

307. *Id.* at 731; *see also* Macedo v. Della Russo, 840 A.2d 238, 242 (N.J. 2004) ("[A]dvertisements by learned professionals in respect to the rendering of professional services are insulated from the [Little FTC Act] but subject to comprehensive regulation by the relevant regulatory bodies and to any common-law remedies that might otherwise apply.").

308. *Averill*, 761 A.2d at 1089 (citing Rousseau v. Eshleman, 519 A.2d 243, 245 (1985)).

"regulatory board."[309] The New Hampshire Legislature has since repealed the relevant language, thus suggesting legislative intent to include professionals within the New Hampshire Little FTC Act.[310]

IV. ANALYSIS

There is little rationale or justification for the hodgepodge of conflicting court interpretations exempting licensed professionals from state Little FTC Acts. The FTC Act and the Sherman Act clearly cover licensed professionals. Additionally, it is clear that most of the state legislatures enacting Little FTC Acts sought to keep their Acts consistent with the FTC and the Acts of other states, united in a single body of precedent. But today, the conduct of a doctor or a lawyer in one state may be actionable under the Little FTC Act, while in another state, the same conduct would not be actionable under an identically worded statute. Even worse, the entire decision may be predicated on whether a judge subjectively determines that the questionable practice was one of entrepreneurialism or of professional judgment.

This particular taxonomy, the accepted practice for judges in Connecticut, Washington, and other states, is fraught with peril. If a law firm pads its bills, is that entrepreneurialism run amuck, or is it a lapse in professional judgment? If a plastic surgeon advertises a procedure to improve one's looks and it fails, is that false advertising akin to rabid entrepreneurialism, or can it be excused as a professional failure outside the scope of the Little FTC Act? If a certified public accountant fails to give one's finances the attention promised in print ads, can the wrong be characterized solely as malpractice, or is it also false advertising?

This still overlooks the procedural issue of whether a client or a patient must believe that the service failed to live up to their expectations or whether it is a proverbial "reasonable man" that must think so. All of this is judge-made law that is beyond the clear language of the Little FTC Acts. No matter how one parses out these distinctions, they fail to provide any consistent or predictable legal standard that can be associated with the rule of law.

Some courts have created exemptions based on the fact that the professions are regulated by regulatory or governing bodies. The New Hampshire Supreme Court, for example, held that

> [t]he statutory exemption to the Act, however, does not require that remedies available to aggrieved consumers under qualifying regulatory schemes be *identical* to those provided in the Act. Rather, it is sufficient that the regulatory scheme protects consumers from fraud and deception in the marketplace . . . . We conclude that the disciplinary measures exercised by

309. *Id.* at 1087; *see* N.H. REV. STAT. ANN. § 358-A:3(I) (2000) (superseded).; Rousseau v. Eshleman, 519 A.2d 243, 245 (N.H. 1986) (interpreting the state Little FTC Act as exempting professionals regulated by a professional board or statute; post *Averill*, the state legislature repealed the relevant statute).

310. N.H. REV. STAT. ANN. § 358-A:3(I) (2005)

> this court protect the public as effectively from deceptive or unfair actions in the marketplace as would double or treble damages under the Act.[311]

But this begs the question. The actual words of the Little FTC Acts suggest two purposes: 1) multiple damage awards as a deterrent against unfair competition or unfair or deceptive acts by providers of goods and services; and 2) allowing direct compensation of alleged victims by permitting private attorney general suits with potential awards of multiple damages.[312] Relegating these matters to a state disciplinary board may, at best, deter future unlawful conduct by punishing unethical professionals. But, under the regulatory scheme envisioned by the New Hampshire Supreme Court and others, the victims remain uncompensated for their injuries. The fact that the professional may have his or her license suspended or revoked is unlikely to be of any solace to someone who has been injured, especially since, for example, disbarment is rarely a permanent ban from the practice of law.[313]

Further, the adequacy of protection assumes that all state regulatory boards are efficient, competent, and fair-minded—something that may not be the case.[314] It also ignores the fact that other remedies, such as malpractice and other tort damages, are increasingly unavailable because of state caps, limitations, and other measures intended to chill excessive and frivolous litigation.[315] Moreover, under alternative causes of action, plaintiffs may face an onerous burden. For example, suing under common law fraud requires proof of intent,[316] and suing for legal malpractice requires proof that the

---

311. *Averill*, 761 A.2d at 1089 (citations omitted).

312. PRIDGEN, *supra* note 92, § 3:2.

313. Donna S. Harkness, *Packaged and Sold: Subjecting Elder Law Practice To Consumer Protection Laws*, 11 J.L. & POL'Y 525, 530 (2003).

314. *See, e.g.*, Keith R. Fisher, *The Higher Calling: Regulation of Lawyers Post-Enron*, 37 U. MICH. J.L. REFORM 1017, 1018 (2004) (discussing the ethical challenges and enforcement issues that lawyers and regulatory boards face in today's society); Michael S. Frisch, *No Stone Left Unturned: The Failure of Attorney Self-Regulation in the District of Columbia*, 18 GEO. J. LEGAL ETHICS 325, 325 (2005) (discussing the problems with self-regulatory boards monitoring attorneys).

315. *See, e.g.*, Geoff Boehm, *Debunking Medical Malpractice Myths: Unraveling the False Premises Behind "Tort Reform,"* 5 YALE J. HEALTH POL'Y, L. & ETHICS 357, 358-62 (2005) (discussing the fact that caps hurt the most severely injured patients); Benjamin H. Davidson et al., *Texas Statutory Caps And Settlement Credits After House Bill 4*, 46 S. TEX. L. REV. 1217, 1218 (2005) ("The recent enactment of House Bill 4 . . . imposed various statutory liability caps on healthcare liability claims . . . .").

316. *See* Harkness, *supra* note 313, at 552; David L. Hudson Jr., *Professional Judgment*, A.B.A. J., July 2003, at 14, 15 ("There are advantages to suing under the consumer protection statute[, because t]he burden of proof is not as onerous as in a common-law fraud claim, where a plaintiff must prove intent."); Albert N. Shelden, *State Consumer Protection Laws and Unfair Competition*, SG076 A.L.I.-A.B.A. 487, 504 (2002) ("Case law clearly indicates that an action predicated on a violation of these statutes is not the same as an action for fraud or common law deceit: intent to deceive, reliance, and damage are not elements of a consumer protection cause of action.").

plaintiff would have prevailed on the underlying case as well as proof that the attorney failed to meet the appropriate standard of care.[317]

In addition, there is an unnecessary and problematic disconnect between antitrust and consumer protection doctrines. While neither doctrine should presume to be binding authority on the other in all instances, it is hard to see how judges considering conduct under a Little FTC Act can describe a case as one of first impression when the relevant statute is taken from the FTC Act, and the United States Supreme Court has found the relevant conduct unlawful under an identically worded statute.[318] The majority of Little FTC Acts ask courts to be guided by the FTC and the actions of other states.[319] Therefore, antitrust doctrines should be persuasive authority in consumer protection matters, particularly those involving licensed professionals.

While some writers tracing the common origins of antitrust and consumer protection have referred to "consumer sovereignty"[320] or have used the metaphor of an "umbrella,"[321] the problem with licensed professionals being treated differently under the two doctrines may be best answered by a "Transitive Principle of Antitrust and Consumer Protection." A transitive relationship is one in which, for example, if A equals B and B equals C, then A equals C.[322] Here, the transitive relationship means that, if the Sherman Act precedent is absorbed into the FTC Act, and FTC Act precedent is absorbed into the Little FTC Acts,[323] then the Sherman Act precedent is necessarily absorbed into the Little FTC Acts. Although apparently never explicitly

---

317. RONALD E. MALLEN & JEFFREY M. SMITH, LEGAL MALPRACTICE, § 32.1 at 128 (2005 ed.).

318. *Compare* FTC v. Ind. Fed'n of Dentists, 476 U.S. 447, 451-52, 465-66 (1986) (holding that withholding a patient's x-rays is a "violation of § 1 of the Sherman Act"); Cal. Dental Ass'n v. FTC 224 F.3d 942, 951-52, 958 (9th Cir. 2000) (holding that restraints on advertising are violations of antitrust law), *with* Rosseau v. Eshleman, 519 A.2d 243, 245 (N.H. 1986) ("In this case of first impression, we must determine whether attorneys in this State are exempt from the application of the consumer protection act.").

319. *See, e.g.*, MASS. GEN. LAWS ANN. ch. 93A, § 2(b) (West 1997).

320. Averitt & Lande, *supra* note 2, at 713

> Antitrust and consumer protection law share a common purpose in that both are intended to facilitate the exercise of consumer sovereignty or effective consumer choice. Consumer sovereignty exists when two fundamental conditions are present. There must be a range of consumer options made possible through competition, and consumers must be able to choose effectively among these options.

Id.

321. Caswell O. Hobbs, Prepared Remarks for the Panel Discussion on "*Under One Unbrella [sic]: Integrating the Competition and Consumer Protection Missions*" at the Federal Trade Commission, 90th Anniversary Syposium [sic], *Antitrust and Consumer Protection—Exploring The Common Ground* (Sept. 22, 2004), *at* http://www.ftc.gov/ftc/history/docs/041123hobbs.pdf (last visited September 6, 2005).

322. *See* Christopher D. Stone, *Should Trees Have Standing? Revisited: How Far Will Law And Morals Reach? A Pluralist Perspective*, 59 S. CAL. L. REV. 1, 92 (1985).

323. At least for those states that have legislatively enacted a requirement of harmonization with the FTC or with the other states that have similar statutes or both.

addressed, many courts adjudicating their Little FTC Act cases refer to the Sherman Act precedent without any explanation, so, in practice, this is not a radical theory.[324]

Out of context, the results of such a transitive analysis of the law might range from wild speculation to merely being dead wrong. But, within the context of antitrust and consumer protection, there is ample support for such a transitive principle, cementing the connection between antitrust and consumer protection doctrines. Antitrust could be and should be used as persuasive authority for consumer protection, at least when faced with a consumer protection case of first impression and little explicit legislative guidance.

The Sherman Act is the heart of antitrust law. The FTC Act is the keystone between antitrust and consumer protection law, governing them both. The FTC Act also includes—at least on the antitrust side—all of the Sherman Act jurisprudence.

The Little FTC Acts contain the key language from the FTC Act; thus, it makes little sense to completely divorce the Little FTC Acts from the rich antitrust history and precedent that led to their creation. Therefore, the Sherman Act jurisprudence must have some bearing on the Little FTC Acts. Given a case involving a licensed professional and a mixed question of antitrust and consumer protection, would the licensed professional be regulated by the Sherman and the FTC Acts, but exempt under a Little FTC Act, which cites the same statutory language?

There have been many cases over the years which have involved both antitrust and consumer protection doctrines.[325] Perhaps none of them have included an issue when antitrust had one set of precedents and consumer protection had another, but if that were to happen, how could the two doctrines possibly be reconciled?

Cases of first impression, aggravated by a gap or vagueness in a statute, require courts to make educated guesses as to what the legislature actually intended.[326] In such situations, judges considering the application of state consumer protection laws to licensed professionals have come to different conclusions, at different times, for different reasons.[327] While this can be written off as a mere wrinkle in America's federal and judicial system of

---

324. *See, e.g.*, Short v. Demopolis, 691 P.2d 163, 168 (Wash. 1984) ("The more recent federal cases stand for the principle that attorneys, as well as other professionals, are not exempt from antitrust laws.").

325. *See, e.g.*, *In re* Pfizer, Inc., 81 F.T.C. 23, 57 (1972).

326. *See* Girardeau A. Spann, *Deconstructing the Legislative Veto*, 68 Minn. L. Rev. 473, 530 (1984).

327. *Compare* Averill v. Cox, 761 A.2d 1083, 1089-90 (N.H. 2000) (stating that holding attorneys exempt from the state consumer protection act is supported by the legislature's silence on the matter); Rosseau v. Eshleman, 519 A.2d 243, 245-46 (N.H. 1986) (holding that the legislature must clearly express a desire to change the interpretation of the consumer protection act in order to change the current holding that attorneys are exempt), *with* Gilmore v. Bradgate Assocs., 604 A.2d 555, 557 (N.H. 1992) (limiting the holding in *Rousseau* to attorneys).

government, it does create unnecessary problems and conflict. Certainly when judges consider for the first time whether licensed professionals should be covered by the state consumer protection law, rules of construction strongly suggest that judges should look to the source of the state law. If that source is the FTC Act, then the answer can be found in the ample and well-reasoned case law finding no special exemption for licensed professionals under the federal antitrust laws.[328]

V. CONCLUSION

Application of the Little FTC Acts to licensed professionals is not only well-reasoned based on decades of precedent, it is consistent with growing public policy concerns. Many states, for example, have supplemented the Little FTC Acts over the past few years to increase the protection of senior citizens, who are frequently the victims of fraud.[329] According to the FTC, "older consumers are more likely to fall victim to certain types of fraud."[330] Kentucky's maximum penalty when the consumer fraud victim is over the age of sixty is five times greater than for younger victims.[331] "Although health care-related complaints are not among the top categories reported to the FTC by older consumers, they nonetheless represent a great concern to the Commission because the consequences of health-related fraud can be so dire."[332] It is almost unfathomable why, if this is such a big concern, doctors (as well as all other licensed professionals) are exempt from some state's Little FTC Acts.

Perhaps the underlying public policy issue behind much of this debate is whether some of the most powerful individuals in our society, lawyers, doctors, and other licensed professionals, have used their privileged positions in society to place themselves above or beyond the law. According to one Washington-based public interest group, "'[t]he idea that lawyers should be above a law that applies to every other business person is patently absurd.'"[333]

---

328. *See supra* Part III.A-B.

329. *See, e.g.*, Harkness, *supra* note 313, at 574-77 ("[S]everal state consumer protection statutes . . . contain special provisions protecting the elderly."); Elaine Markowitz, *Lawyer's Mission is Helping*, TAMPA TRIBUNE, Dec. 4, 2004, at 11 ("The lawyers . . . help people with consumer issues, particularly the elderly, who are often victims of fraud and identity theft.").

330. *Identifying and Fighting Consumer Fraud Against Older Americans, Testimony to the Senate Special Committee on Aging* 2 (2005) (prepared statement of Lois C. Greisman, Associate Director of the Division of Planning and Information, Bureau of Consumer Protection, U.S. Federal Trade Commission), *available at* http://www.ftc.gov/os/testimony/050727confraudolder.pdf (last visited Mar. 20, 2006).

331. KY. REV. STAT. ANN. § 367.990(2) (LexisNexis 2002).

332. *Identifying and Fighting Consumer Fraud Against Older Americans, supra* note 330, at 8.

333. Richard B. Schmitt, *Widow's Fight Tests New Way To Sue Lawyers*, HALT, Nov. 13, 2005 (quoting James Turner, executive director of HALT), *at* http://www.halt.org/about_halt/in_the_news/widow_sues_lawyer.php (last visited Mar. 20, 2006).

The comment was made in response to an effort on the part of the 35,000 member Illinois Bar Association to persuade the court to exempt attorneys from the Illinois Little FTC Act[334]—an effort that was eventually successful.[335]

There is no sound basis in law, history, precedent or public policy for exempting licensed professionals from the reach of the Little FTC Acts. Moreover, when courts choose to do so, the distinctions are usually based on subjective criteria, such as entrepreneurialism, that do not generally stand up to scrutiny. Finally, society as a whole must ask whether these exemptions are created solely to favor those who are best capable of lobbying for themselves; with this in mind, there is little persuasive argument favoring an exemption of licensed professionals.

334. *Id.*

335. Cripe v. Leiter, 703 N.E.2d 100, 106 (Ill. 1998).

THE ROLE OF RETRIBUTIVE JUSTICE IN THE COMMON LAW OF TORTS: A DESCRIPTIVE THEORY

RONEN PERRY*

Justitia est constans et perpetua voluntas jus suum cuique tribuens[1]

I. INTRODUCTION

Retribution, or just desert, is usually perceived as one of the most prominent theoretical foundations of *criminal liability*.[2] Tort jurisprudence does not appear to be the natural habitat for retributive concerns. The principle of just desert focuses on a single person and inflicts upon him a sanction whose severity is determined solely by the gravity of his wrongdoing.[3] In contrast, tort law focuses on bipolar interactions. It links the wrongdoer directly to the victim of his wrong and obliges the former to compensate the latter for her injuries if certain conditions are met. The sanction is determined by the specific plaintiff's loss, not by the gravity of the defendant's wrong. Given its bipolar structure and rectificatory function, tort liability is technically inconsistent with the notion of retributive justice. This may explain why just desert is usually neglected in the theoretical analysis of tort law.

A few words must be said about existing theories of tort law. In the second edition of *Torts and Compensation*, Professor Dobbs presented the current debate in tort scholarship as resting mainly on this question: "Should judges decide cases on the basis of corrective justice with a view to righting wrongs and doing justice between the parties? Or should they decide cases according to

---

* Faculty of Law, University of Haifa; LL.D. summa cum laude (Hebrew University, 2000), LL.M. studies completed with distinction (Hebrew University, 1997), LL.B. magna cum laude (Tel Aviv University, 1996). I would like to thank the Honorable Izhak Englard (of the Supreme Court of Israel), Israel Gilead, Ariel Porat, and the editors of the Tennessee Law Review for their helpful comments.

1. "Justice is the constant and perpetual wish to render every one his due." I THE INSTITUTES OF JUSTINIAN, D.i.1.10. (Thomas C. Sandars trans., Longmans, Green and Co. 1934) (The four books of The Institutes of Justinian were originally issued in 535 C.E.).

2. Paul Butler, *Retribution, For Liberals*, 46 UCLA L. REV. 1873, 1874 (1999); Michele Cotton, *Back with a Vengeance: The Resilience of Retribution as an Articulated Purpose of Criminal Punishment*, 37 AM. CRIM. L. REV. 1313, 1315 (2000); Gary T. Schwartz, *Mixed Theories of Tort Law: Affirming Both Deterrence and Corrective Justice*, 75 TEX. L. REV. 1801, 1811-12 (1997) [hereinafter Schwartz, *Mixed Theories*]. Other popular justifications for criminal liability are deterrence, rehabilitation or treatment, and incapacitation. Cotton, *supra*, at 1315.

3. See discussion *infra* Parts II.A, II.B.

what is good for society as a whole?"[4] This statement may give the reader a rudimentary understanding of what appears to be the most fundamental controversy in tort theory.[5] In a way, this is a misstatement of the true nature of the theoretical dispute. Corrective justice is a *form* of justice and not a *substantive* moral standard. It determines *how* justice is done, but not *when* it should be done.[6] Corrective justice has been defined as an arithmetical rectification of a wrong committed through a bilateral interaction.[7] To determine what a wrong is, one needs a substantive moral theory, be it deontological or consequentialist. Clearly, an economist may say that a person commits a "wrong" when she acts in an inefficient way.[8] In other words, corrective justice does not necessarily exclude welfare maximization. The latter is a substantive moral commitment, whereas the former is mere form. Consequently, describing the current debate in tort scholarship as a conflict between corrective justice theory and economic theory is somewhat inaccurate, but understandable.

Corrective justice theorists usually fuse Aristotle's notion of corrective justice with a substantive non-consequentialist moral standard, most notably Kant's principle of right,[9] thereby depriving corrective justice of its true meaning as a mere mathematical form. On the other hand, law-and-economics theorists focus on substance and seldom use Aristotle's notions of justice.[10] So it may look as if corrective justice and welfare maximization are in dispute, although the true conflict is a substantive moral one. In any event, there seem

---

4. DAN B. DOBBS, TORTS AND COMPENSATION 840-41 (2d ed. 1993); *cf.* Schwartz, *Mixed Theories*, *supra* note 2, at 1801 ("Currently there are two major camps of tort scholars. One understands tort liability as an instrument aimed largely at the goal of deterrence, commonly explained within the framework of economics. The other looks at tort law as a way of achieving corrective justice between the parties.").

5. Dobbs is using a normative rhetoric (*should* judges decide), but his characterization of the current theoretical debate is equally applicable on the descriptive level.

6. *Cf.* Louis Kaplow & Steven Shavell, *Fairness Versus Welfare*, 114 HARV. L. REV. 961, 1046 (2001) ("[C]orrective justice is incomplete in that one must look elsewhere for a substantive theory of what counts as wrongful injury.").

7. ERNEST J. WEINRIB, THE IDEA OF PRIVATE LAW 62-66 (1995) (explaining Aristotle's corrective justice).

8. Richard A. Posner, *The Concept of Corrective Justice in Recent Theories of Tort Law*, 10 J. LEGAL STUD. 187, 201 (1981). Posner noted that "Once the concept of corrective justice is given its correct Aristotelian meaning, it becomes possible to show that it is not only compatible with, but required by, the economic theory of law. In that theory, law is a means of bringing about an efficient (in the sense of wealth maximizing) allocation of resources . . . ."

9. WEINRIB, *supra* note 7, at 80-83, 114-44; Richard W. Wright, *Right, Justice and Tort Law*, *in* PHILOSOPHICAL FOUNDATIONS OF TORT LAW 159, 160-71 (David G. Owen ed., 1995) [hereinafter Wright, *Right*]; Richard W. Wright, *The Standards of Care in Negligence Law*, *in* PHILOSOPHICAL FOUNDATIONS OF TORT LAW 249, 255-63 (David G. Owen ed., 1995). Under the Kantian concept of right, "all that is in question is . . . whether the action of one of the two parties can be united with the freedom of the other in accordance with a universal law." WEINRIB, *supra* note 7, at 95.

10. *See generally*, STEVEN SHAVELL, ECONOMIC ANALYSIS OF ACCIDENT LAW (1987) (developing a comprehensive economic theory of tort law without referring to Aristotle).

to be only two major contestants on the theoretical battleground: corrective justice and welfare maximization.

As is clearly evident from the title, this article attempts to fill a gap in existing tort theory. It transcends the current theoretical debate and asserts that the development of tort doctrine has been influenced or inspired – to a limited extent – by the notion of retributive justice. My first, rather modest, goal is to elucidate the concept of retributive justice as an independent *form* of justice. My second, more ambitious, goal is to offer a descriptive theory of the role of retributive justice in the common law of torts. I contend that this notion is entrenched in tort law, although it operates only in fairly limited circumstances.

Part II defines retributive justice, and explains what kind of social inequity it is supposed to set right, and how it operates to remove this inequity. Further, it defends the view that retributive justice is an independent *form* of justice, not a subset of one of the two classical Aristotelian forms. Finally, it surveys various factors that may be taken into account in determining the moral gravity of human conduct, the elemental input in any retributive operation.

Part III is the cornerstone of my thesis. It explains why retributive justice cannot be regarded as the principal rationale of tort liability and argues that it nevertheless plays a limited, but noteworthy, role in tort jurisprudence. The exceptional circumstances, in which retributive justice is applied, may be subsumed under two general paradigms: the prevention of *abominable* disproportion paradigm and the preservation of criminal justice paradigm. In cases of the first type, tort law strives to prevent an acute imbalance between the severity of the sanction imposed on one of the litigating parties and the gravity of her conduct. In cases of the second type, tort law attempts to vindicate criminal justice as prescribed by criminal law, thereby advancing the goals of the latter.

Because the tort process links two persons and because there are two possible deviations from the principle of retributive justice (over-punishment and under-punishment), this principle may be used to justify four types of argument in tort law: (1) liability should be *limited*, or excluded, to prevent over-punishment of the *injurer*; (2) liability should be *limited*, or excluded, to prevent under-punishment of the *victim*; (3) liability should be *expanded* to prevent under-punishment of the *injurer*, and (4) liability should be *expanded* to prevent over-punishment of the *victim*. The first and second possibilities are discussed in Part IV, the third and fourth in Part V.

II. THE NOTION OF RETRIBUTIVE JUSTICE

A. Defining Retribution

The notion of retribution has been one of the most prominent justifications for criminal punishment for centuries.[11] In his magnum opus, *Metaphysische*

11. *See* sources cited *supra* note 2.

Anfangsgründe der Rechtslehre, Immanuel Kant argued that this was the only possible justification for punishing lawbreakers.[12] In his translated words:

> Judicial punishment can never be used merely as a means to promote some other good for the criminal himself or for civil society, but instead it must in all cases be imposed on him only on the ground that he has committed a crime . . . . He must first be found to be deserving of punishment before any consideration is given to the utility of this punishment for himself or for his fellow citizens.[13]

Retribution may thus be defined as imposing a sanction that corresponds to individual moral desert.[14] The wrongdoer deserves to be punished on account of her wrongful conduct, and ought to be punished fairly regardless of the consequences of her punishment.[15] Why do wrongdoers *deserve* to be punished? One possible reason is that retribution communicates the value of the victim by disproving a false claim of superiority implied by a wrongful act.[16] The wrongdoer treated her victim as inferior, and the sanction nullifies any explicit, or implicit, assertion of the former's superiority. This explanation does not seem very persuasive. The law punishes wrongdoers even when the wrong shows no affront to the victim's value, and it can hardly be said that doing so is inherently unfair in the retributive sense.

A more convincing reason involves the notion of reciprocity. Every person enjoys the fact that all people abide by legal rules. Every person must obey these rules in exchange for the concessions made by others who conform to the same rules. If a person does not abide by a specific rule, societal balance is undermined. Punishment is the price she is compelled to pay for the concessions made by law-abiding citizens, given that she did not pay for them

---

12. IMMANUEL KANT, METAPHYSICAL ELEMENTS OF JUSTICE (John Ladd trans., Hackett Publishing Co., 2nd ed. 1999) (1785).

13. *Id.* at 138.

14. *See* George P. Fletcher, *The Place of Victims in the Theory of Retribution*, 3 BUFF. CRIM. L. REV. 51, 52 (1999) ("[B]y the use of the easily misunderstood term 'retributive', I simply mean imposing punishment because it is deserved on the basis of having committed a crime . . . . [R]etributivism is a jealous theory in the sense that whatever the beneficial side-effects of punishment, if it is not deserved it cannot possibly be justified.").

15. Retributive justice is, therefore, *retrospective*, in that it looks backward to the particular wrongdoing, not onward to the consequences of the sanction.

16. *See* Jean Hampton, *The Retributive Idea*, *in* FORGIVENESS AND MERCY 111, 122-47 (Jeffrie G. Murphy & Jean Hampton eds., 1988) (explaining retribution as a defense of human value); Jean Hampton, *Correcting Harms Versus Righting Wrongs: The Goal of Retribution*, 39 UCLA L. REV. 1659, 1686 (1992) [hereinafter Hampton, *The Goal of Retribution*]. Hampton defines retribution as

> [A] response to a wrong that is intended to vindicate the value of the victim denied by the wrongdoer's action through the construction of an event that not only repudiates the action's message of superiority over the victim but does so in a way that confirms them as equal by virtue of their humanity.

Id. See also Andrew E. Taslitz, *The Inadequacies of Civil Society: Law's Complementary Role in Regulating Harmful Speech*, 1 MARGINS 305, 314-15 (2001).

in the usual way, namely by conforming to the same rules.[17]

A third, and more plausible, reason is that wrongful conduct may have two aspects that ought to be dealt with: a private aspect, represented by the *harm* caused to the immediate victim of the wrong, and a public aspect, represented by the aggregate outrage, dissatisfaction, and loss of confidence incurred by all members of society due to the occurrence of the wrong. The private aspect is the concern of corrective justice. The wrongdoer is obliged to rectify the private harm caused by her conduct. However, the public aspect of the wrong must also be addressed. The wrongdoer deserves to pay for it somehow, in order to restore the *status quo ante* in full and annul the wrong. Retribution is aimed at rectifying the public aspect of wrongful conduct committed by individuals.[18]

Retributive justice does not require that the sanction be identical to the wrong committed, unlike the ancient *lex talionis;*[19] it merely insists on *proportionality* between the severity of the sanction and the gravity of the wrong.[20] The sanction must be fair in light of the conceptual and concrete features of the wrong for which it is imposed.[21] The fairness of a legal sanction is determined by two complementary principles: cardinal, or absolute, proportionality and ordinal, or relative, proportionality.[22]

---

17. *See* JEFFRIE G. MURPHY, KANT: THE PHILOSOPHY OF RIGHT 142-43 (1970) ("If the law is to remain just, it is important to guarantee that those who disobey it will not gain an unfair advantage over those who do obey voluntarily.")

18. Retribution is different from satisfying the victim's feeling of indignation. The latter can be achieved, at least partially, through bilateral private litigation, and is therefore within the corrective justice domain. Walter J. Blum & Harry Kalven, Jr., *The Empty Cabinet of Dr. Calabresi*: Auto Accidents and General Deterrence, 34 U. CHI. L. REV. 239, 268-69 (1967) ("In large part corrective justice is concerned . . . with satisfying the victim's feeling of indignation. If the victim [cannot sue the injurer], he will not get the satisfaction of seeing his wrong righted." (footnote omitted)).

19. "The law of retaliation, which requires the infliction upon a wrongdoer of the same injury which he has caused to another." BLACK'S LAW DICTIONARY 913 (6th ed. 1990).

20. Peter Cane, *Retribution, Proportionality, and Moral Luck in Tort Law*, *in* THE LAW OF OBLIGATIONS 141, 143, 160-61 (Peter Cane & Jane Stepleton eds., 1998); Tony Honoré, *The Morality of Tort Law - Questions and Answers*, *in* PHILOSOPHICAL FOUNDATIONS OF TORT LAW 73, 87 (David G. Owen ed., 1995) [hereinafter Honoré, *The Morality of Tort Law*] (retribution *requires* imposition of a sanction that is in proportion to the moral gravity of the misconduct, and *forbids* imposition of a sanction that is out of proportion to the gravity of the misconduct); TONY HONORÉ, RESPONSIBILITY AND FAULT 13, 83-84, 92, 123, 138 (1999) [hereinafter HONORÉ, RESPONSIBILITY] (recognizing that the principle of retribution requires that a penalty should not be disproportionate to the moral gravity of the offence); Hampton, *The Goal of Retribution*, *supra* note 16, at 1690; Jeffrie G. Murphy, *Does Kant Have a Theory of Punishment?*, 87 COLUM. L. REV. 509, 530-32 (1987); Taslitz, *supra* note 16, at 335; Note, *Punitive Damages and Libel Law*, 98 HARV. L. REV. 847, 851 (1985) ("Fairness demands that the punishment be proportionate to the severity of the act.").

21. Honoré, *The Morality of Tort*, *supra* note 20, at 86-87. Honoré discusses how retribution requires "a rough correlation between the type of fault or conduct and the weight of the punishment imposed." *Id.* at 87.

22. ANDREW VON HIRSCH, CENSURE AND SANCTIONS 29-46 (1993); ANDREW VON HIRSCH,

According to the principle of cardinal proportionality, the sanction should not be too harsh or too lenient with respect to the *absolute* gravity of the wrong committed.[23] For example, it seems unfair to impose a life sentence on a person who did not pay for parking; similarly, it seems unfair to impose a small fine on a cold-blooded murderer. According to the principle of ordinal proportionality, the sanction imposed for a certain wrong must reflect the *relative* gravity of the wrong: if wrong X is more serious than wrong Y, the sanction for wrong X must be more severe than the sanction for wrong Y, and vice versa.[24] For example, the punishment for murder must always be more severe than the punishment for non-payment for parking.

The principle of cardinal proportionality sets the upper and lower limits of the possible sanction in a given society regardless of the relative gravity of the wrong. The principle of ordinal proportionality narrows those boundaries, so that the order of harshness of actual sanctions will correspond to the order of gravity of the given wrongs. Suppose, for example, that in a certain society there are ten possible methods of punishment: (1) mere reproach; (2) small fine; (3) medium fine; (4) large fine; (5) short term imprisonment (6-12 months); (6) medium term imprisonment (1-10 years); (7) long term imprisonment (10-25 years); (8) life sentence; (9) life sentence with penal servitude; (10) capital punishment. Suppose further that in the same society four offenses frequently occur: (1) traffic misdemeanors; (2) larcenies; (3) armed robberies; (4) murders.

Applying the principle of cardinal proportionality may lead to the following conclusions. A person who commits a traffic misdemeanor deserves a punishment in the range of (1)-(5); a person who commits larceny deserves a punishment in the range of (4)-(7);[25] a person who commits armed robbery deserves a punishment in the range of (6)-(9);[26] and a person who commits murder deserves a punishment in the range of (7)-(10). Yet, according to the principle of ordinal proportionality, the sanction imposed for a given offense must not be similar to, or less than, the sanction imposed for a less serious offense. If sanction (5) is imposed for a traffic misdemeanor, sanction (4) or (5) cannot be imposed for larceny; if sanction (7) is imposed for larceny, sanction (6) or (7) cannot be imposed for armed robbery; and if sanction (9) is imposed for armed robbery, sanction (8) or (9) cannot be imposed for murder, so only capital punishment is left.[27]

PAST OR FUTURE CRIMES: DESERVEDNESS AND DANGEROUSNESS IN THE SENTENCING OF CRIMINALS 40-46 (1985); Cane, *supra* note 20, at 143, 161; Dorsey D. Ellis, Jr., *Fairness and Efficiency in the Law of Punitive Damages*, 56 S. CAL. L. REV. 1, 6-7 (1982).

23. Cane, *supra* note 20, at 143.

24. *Id.*

25. *See* Solem v. Helm, 463 U.S. 277, 303 (1983) (concluding that life imprisonment is an excessive punishment for a nonviolent felony).

26. *See* Coker v. Georgia, 433 U.S. 584, 592 (1977) (concluding that the death penalty is an excessive punishment for rape).

27. For further discussion of proportionality see *supra* notes 20-24. *Cf.* Youngjae Lee, *The Constitutional Right against Excessive Punishment*, 91 VA. L. REV. 677, 683-87 (2005) (contending that the "Eighth Amendment [prohibition] on excessive punishment should be

B. Retribution as a Third Form of Justice

Aristotle distinguished between two forms of justice: corrective, or rectificatory, and distributive.[28] The first is defined as a rectification of harm, wrongfully caused by one person to another, by means of a direct transfer of resources from the injurer to the victim. It assumes an initial arithmetical equality between the parties and aims at restoring that equality whenever it is undermined by a wrongful act.[29] Distributive justice is defined as an allocation of resources or burdens in accordance with the relative merit of each participant, which is determined by a general political criterion. It aligns each person's share of the resources, or burdens, with her merit. It strives for "geometric" equality, that is, equality of ratios. If Jack's merit is *a* and Jill's merit is *b*, and if Jack's share is *c* and Jill's is *d*, distributive justice only prescribes that $a/b=c/d$, and $a/c=b/d$.[30]

After discussing the two forms of justice, Aristotle remarks that the Pythagoreans defined justice "without qualification" as reciprocity.[31] The Pythagorean idea of reciprocity denotes *lex talionis*, the law of identical and direct retaliation, which posits "should a man suffer what he did, right justice would be done."[32] Aristotle contends that this *general* notion of reciprocity fits neither distributive nor corrective justice,[33] although he does not thoroughly explain his assertion.[34] Still, he admits that a qualified application of a refined version of the Pythagorean idea may be deemed just. He maintains that in associations for exchange (i.e. voluntary transactions) reciprocity in accordance with a proportion (i.e. proportionate return) and not on the basis of precisely equal return—holds men together.[35] In his subsequent discussion of proportionate return Aristotle apparently reduces the notion of reciprocity to a principle of fair bargaining, which may be termed commutative justice.[36] It

---

understood as a side constraint" that embodies the principles of ordinal and cardinal proportionality).

28. ARISTOTLE, NICOMACHEAN ETHICS 1130b30-33 (David Ross trans., rev. ed. 1980). Although the Ross translation is highly esteemed in academic circles, it does not provide precise references to Aristotle's original work. The references were consequently taken from ST. THOMAS AQUINAS, COMMENTARY ON ARISTOTLE'S NICOMACHEAN ETHICS (C.I. Litzinger trans., 1964) [hereinafter AQUINAS, COMMENTARY].

29. *Id.* at 1131b25-1132b20.

30. *Id.* at 1131a15-b24. For an interesting discussion of the two types of justice see THOMAS AQUINAS, SUMMA THEOLOGICA, pt. II-II, q. 61 (Fathers of the Eng. Dominican Province trans., 2nd ed. 1920) (1273) [hereinafter AQUINAS, SUMMA].

31. ARISTOTLE, *supra* note 28, at 1132b21-23.

32. *Id.* at 1132b24-27.

33. *Id.* at 1132b23-27.

34. *Id.* at 1132b28-31.

35. *Id.* at 1132b31-1133a5. In other words, "justice" in commercial transactions means that considerations must be proportionate, but not identical.

36. Note, however, that Aristotle's discussion of reciprocity has always presented problems of interpretation. There has never been a consensus as to what reciprocity is or how it is related to the overall topic of justice. *See* Gabriel Danzig, *The Political Character of*

may be argued that this principle conforms to corrective justice, in its broad sense, as it shows how to rectify the social imbalance created by one-sided performance in voluntary bilateral transactions.[37] However, Aristotle seems to recognize that the idea of proportionate return is also applicable to involuntary transactions, wherein one person wrongs another. This may be inferred from the following paragraph:

For it is by proportionate requital that the city holds together. Men seek to return either evil for evil—and if they cannot do so, think their position mere slavery—or good for good—and if they cannot do so there is no exchange, but it is by exchange that they hold together.[38]

The first sentence restates Aristotle's divergence from the ancient *lex talionis*: reciprocity requires proportionate requital, not an identical response. The second sentence consists of two parts; the second part applies the principle of proportionate requital to voluntary exchange. The first part seems to apply the same principle to wrongdoing: people seek to return evil for evil, yet the response ought to be proportionate to the initial evil. This seems to be an underdeveloped retributive intuition.[39] Aristotle does not elaborate on the subject.

According to one view, corrective justice and distributive justice, as defined by Aristotle, exhaust the possible *modes* of justice.[40] Consequently, retributive justice must be either a mere "subset of corrective justice"[41] or a manifestation of distributive justice. I believe that this assertion is flawed, and that Aristotle's implicit intuition is correct. A third, retributive, form of justice exists that is incompatible with either corrective or distributive justice.[42]

In my view, retributive justice is not a species of corrective justice. Corrective justice, unlike retributive justice, ignores the gravity of a person's conduct if it has caused no harm to her legal adversary. Causation of harm is central to the application of corrective justice. There can be no rectification of harm in the absence of harm. Consequently, if both A and B expose C to identical unreasonable risks, but only the risk created by A materializes, corrective justice dictates that C may sue A, but not B.[43] A retributive model

---

Aristotelian Reciprocity, 95 CLASSICAL PHILOLOGY 399, 401-04 (2000).

37. *Id.* at 408-11.

38. ARISTOTLE, *supra* note 28, at 1132b33-1133a5.

39. *See* Jeremy Waldron, *Does Law Promise Justice?*, 17 GA. ST. U. L. REV. 759, 772 (2001) (observing that retributive justice is recognized as a separate form of justice in the Nicomachean Ethics).

40. Wright, *Right*, *supra* note 9, at 174.

41. *Id.* at 175. This view might find support in AQUINAS, COMMENTARY, *supra* note 28, at 308-09.

42. Anglo-American scholars frequently refer to retributive justice as a distinct form, but they do not explain why it is independent of the two Aristotelian forms. *See* Fletcher, *supra* note 14, at 58 (observing that "[r]etributive justice combines features of both corrective and distributive justice."); Paul A. LeBel & Richard C. Ausness, *Toward Justice in Tobacco Policymaking: A Critique of Hanson and Logue and an Alternative Approach to the Costs of Cigarettes*, 33 GA. L. REV. 693, 699, 772, 775-76 (1999).

43. WEINRIB, *supra* note 7, at 155-56.

would not tolerate B's impunity. The causation-of-harm requirement in most perceptions of corrective justice makes it inherently incompatible with retributive justice.

It may be argued that retributive justice is nonetheless a species of corrective justice. This may be because retributive justice rectifies either the dignitary injury caused to the victim, through the use of punitive damages, or the non-discrete injury to the dignity and security of each and every member of society, through the imposition of a criminal sanction.[44] So, whatever we may call retribution is, in fact, rectification of harm. Unfortunately, this assertion fails for two reasons. First, tort law overtly redresses intangible injuries. When courts wish to compensate for such losses they do not try to conceal their intention.[45] Therefore, when they claim that a certain sanction is not intended to compensate for harm, as is the case with punitive damages,[46] there is no reason not to believe them.

Second, rectifying a private harm ensuing from a given wrong is different from a retributive attempt to annul its public aspect. Once it is undisputed that a wrong has been committed, corrective justice and retributive justice function differently. Corrective justice *isolates* the wrongdoer and the victim, and *obliges* the former to repair the latter's harm directly.[47] Both the isolation of the bipolar relationship and the obligation directly to rectify a given harm are missing in retributive justice. Retribution is concerned with the aggregate societal outrage, dissatisfaction, and loss of confidence ensuing from the wrongful conduct, and not with the implications of the wrong within a bipolar relationship.[48] It does not oblige the wrongdoer to repair any personal harm directly.

In addition, retribution is not a species of distributive justice. Retribution focuses on the moral desert of a single person, and does not distribute a benefit or a burden among two or more persons. One may argue that retribution is in fact a distribution of sanctions among wrongdoers according to the gravity of their wrongs; however, there is a fundamental difference between a just allocation of sanctions according to the principle of retribution and a just distribution in Aristotelian terminology. Distributive justice deals with the allocation of *certain* benefits or burdens among *certain* persons. If one of the participants got too much, there is at least one other participant who received

---

44. Wright, *Right*, *supra* note 9, at 175.

45. *See* RESTATEMENT (SECOND) OF TORTS § 46(1) (1965) (stating that a person is liable for severe emotional distress caused by his outrageous conduct).

46. *See infra* Section V.A.1.

47. Aristotle himself clearly defines corrective justice as "justice in transactions between man and man" (ARISTOTLE, *supra* note 28, at 1130b33-1131a3). This point is emphasized by AQUINAS, SUMMA, *supra* note 30, pt. II-II, q. 61, art. 3: "commutative justice directs commutations that can take place between two persons."

48. In my view, applying the notion of retributive justice cannot be deemed as vindicating an independent state interest but as an attempt to vindicate the aggregate interest of law abiding citizens. Therefore, it cannot be said that retributive justice is a particular application of corrective justice to the relationship between the state and its mischievous citizen.

too little. Amending a distributive injustice requires a transfer from those who received too much to those who received too little. Conversely, in a retributive allocation of sanctions the class of persons who may deserve punishment for wrongful conduct is indeterminate, and there is no given burden that has to be distributed. Each wrongdoer must get due punishment; an unfair punishment in one case does not swell or shrink the pool of sanctions available for other cases. The fact that one person has been punished too severely or too leniently does not mean that it will be impossible to impose just sanctions on others. Amending an injustice does not require a transfer.

In sum, retributive justice is not an offshoot of either corrective or distributive justice; it is an independent form of justice.[49] It does not directly rectify harms caused through bilateral interactions or distribute a certain burden among certain people. Instead, retributive justice focuses on a single person, namely the wrongdoer, and makes her suffer due to the public disapproval of her conduct.

C. The Substance Behind the Form

The content of the wrong, for which a punishment is deserved, is not consensual, nor does it have to be. Retribution is a *form* of justice, an apparatus, and not an independent *substantive* moral standard.[50] It determines how justice should be done whenever a wrong is committed. The definition of wrong is left to legal philosophers. In practice, courts and scholars tend to evaluate the severity of human conduct in light of some or all of the following features, which are derived from both consequentialist and deontological theories of law.[51]

The foremost feature is the magnitude of the injury expected as a result of the conduct. This feature is determined by three factors:

1. The *value* of the interest that was put at risk by the given conduct. For example, threatening a person's life is generally more acute than jeopardizing her property, or a purely financial interest.[52]

---

49. HONORÉ, RESPONSIBILITY, *supra* note 20, at 7, 13 (enumerating three "elements of justice"—corrective, distributive and retributive); Lea Brilmayer, *International Justice and International Law*, 98 W. VA. L. Rev. 611, 615 (1996) ("The basic conceptions of international justice include retributive justice, corrective justice, and distributive justice.").

50. For this reason I do not find Kaplow and Shavell's criticism of the idea of retribution very appealing. *See* Kaplow & Shavell, *supra* note 6, at 972, 1785-86 ("Retributive theorists assert . . . that punishment should follow automatically from the commission of a wrongful act, but they fail to offer a theory of which acts are wrongful . . . .").

51. For a brief explanation of the deontology or consequentialism dichotomy, see Heidi M. Hurd, *The Deontology of Negligence*, 76 B.U. L. REV. 249, 252-54 (1996) ("Consequentialists are committed to the claim that wrongdoing consists in failing to maximize good consequences and/or minimize bad consequences . . . . According to deontologists . . . what is morally wrong is exclusively a function of an agent's violation of [agent-relative maxims that impose obligations or grant permissions].").

52. Cane, *supra* note 20, at 147. *See* State Farm Mut. Auto. Ins. Co. v. Campbell, 538

2. The prospective *change* in the condition of the interest that was put at risk. Causing a small temporary bodily harm may not be as severe as depriving a person of her entire property. Special vulnerability of the victim may influence the prospective change in the state of her interest.[53] For example, abusing a person may have a more serious impact on her mental health if she is still a small child.
3. The *probability* of the injury. Clearly, exposing a person to a 1% chance of physical injury is not as severe as exposing her to a 100% chance of the same injury (e.g. by assault).

A second, and important, factor considered in determining the gravity of the conduct is the level of the actor's awareness of the risk created by her conduct and of her willingness to cause harm. A person who was unaware of the risk created by her conduct does not deserve the same sanction as a person who was aware of the risk created, and neither of these deserves the same sanction as one who actually intended to cause harm.[54]

Other relevant features bearing on the gravity of the wrongdoer's conduct include: the existence or non-existence of the victim's consent;[55] the repetitive or isolated nature of the conduct and its duration;[56] and the wrongdoer's wealth (the wealthier he is, the more severe monetary sanction is required to give him

---

U.S. 408, 419 (2003) ("We have instructed courts to determine the reprehensibility of a defendant by considering [among other factors] whether: the harm caused was physical as opposed to economic . . . .").

53. *Campbell*, 538 U.S. at 419 (noting that a victim's financial vulnerability is relevant in determining reprehensibility).

54. Norman v. Massachusetts Bay Transp. Auth., 592 N.E.2d 139, 141 (Mass. 1988) ("Society is rightly less concerned about the burden placed on an intentional wrongdoer than about the burden placed on one who has been merely negligent."); Payton v. Abbott Labs, 437 N.E.2d 171, 176 (Mass. 1982)

> [T]he retributive function of imposing tort liability is served by allowing recovery for emotional distress, without proof of physical harm, where a defendant's conduct was either intentional or reckless. Where a defendant was only negligent, his fault is not so great as to require him to compensate the plaintiff for a purely mental disturbance.

Id.; ALAN CALNAN, JUSTICE AND TORT LAW 114 (1997) (explaining that the retributive inclination heavily depends on the wrongdoer's state of mind); Honoré, *The Morality of Tort*, *supra* note 20, at 86-87; Cass R. Sunstein et al., *Assessing Punitive Damages (with Notes on Cognition and Valuation in Law)*, 107 YALE L.J. 2071, 2085-86 (1998) (arguing that societal outrage with certain behaviors may be expressed through retributive punishments); Taslitz, *supra* note 16, at 335-36 (stating that a wrongdoer's intentions can temper, or heighten, the sanction).

55. Martin A. Kotler, *Utility, Autonomy and Motive: A Descriptive Model of the Development of Tort Doctrine*, 58 U. CIN. L. REV. 1231, 1248-54 (1990).

56. *Campbell*, 538 U.S. at 419 (stating that whether the conduct involved repeated actions or was an isolated incident is relevant in determining reprehensibility); Pac. Mut. Life Ins. Co. v. Haslip, 499 U.S. 1, 21 (1991) (stating that the "existence and frequency of similar past conduct" is relevant in determining the magnitude of punitive damages).

his due).[57] Utilitarian theorists would add the cost of precautions needed to prevent the injury, or reduce its probability, and the level of social utility of the risk creating conduct.[58] The foregoing factors are used only to determine the gravity of the wrong committed. They are considered within the substantive moral evaluation of the conduct. When this preliminary issue is resolved, retributive justice calls for the imposition of a sanction that roughly corresponds to the severity of the conduct.

Regardless of any possible theoretical dispute about the significance of each of the foregoing factors, empirical studies show that ordinary people evaluate the severity of various conducts similarly.[59] Not only do they rank the severity of various cases similarly, but they also have similar judgments of the absolute severity of discrete cases on a given scale.[60]

III. THE LIMITED EXPLANATORY POWER OF RETRIBUTIVE JUSTICE IN TORTS

A. The Fundamental Flaws of Monistic Theories

One may argue that the main purpose of the law of torts is to get even with wrongdoers. Not because society wishes to deter future injurers, but because wrongdoers deserve to be penalized. The most prominent advocate of a retributive theory of tort law is probably Professor Martin Kotler, who states:

> [M]uch of the development of tort doctrine can be understood in terms of penalties . . . . [N]ot only the decision of individual cases but also the development of tort doctrine as a whole can be seen as an attempt to punish conduct which violates certain core values that comprise the underlying basis of moral intuition. Punishment in this context is not a means of accomplishing some other goal - efficient cost allocation or accident reduction, for example - but a means of exacting revenge or retribution.[61]

Although most of Kotler's work is dedicated to defining the wrong for which a sanction is due, the main idea is simple: tort law is a retributive mechanism. With respect, I disagree. The notion of retribution may play a significant role in the laymen's understanding of tort law;[62] perhaps this

---

57. Imposing a $1,000 fine on a hard working proletarian may be enough as punishment for accidentally injuring the property of another, but it will not be enough if the injurer is a very wealthy man who will not feel the loss of $1,000. This conclusion derives from the subjective nature of retributive punishment and the diminishing utility of wealth. The community wants to make the injurer suffer in a manner proportional to the wrong committed. To do so we need to know the effect of any sanction on his well-being, and, to assess the subjective impact of the sanction, we need to know how wealthy he is.

58. Kotler, *supra* note 55, at 1244-48.

59. Sunstein, et al., *supra* note 54, at 2098.

60. *Id.*

61. Kotler, *supra* note 55, at 1232.

62. *Id.* at 1233 ("[E]xperience indicates that punishment of wrongdoers is the dominant

understanding has even been endorsed by one or two judicial opinions. [63] But it is a total misconception of tort law from a theoretical standpoint.

First, the law of negligence, which is currently the most significant division of tort law, does not penalize wrongful conduct unless damage ensues; whereas, from a retributive perspective, wrongful conduct must yield the same sanction regardless of the fortuitous occurrence of harm. Whether harm occurs or not is a fortuity that does not alter the gravity of the conduct; hence it should have no effect on the severity of the sanction.[64] There may be only a single attempt to explain why tort liability is retributively just despite its dependence on the occurrence of harm,[65] but it seems to me unconvincing.[66]

Second, tort law often imposes liability for conduct that cannot be deemed morally wrong. This is done especially under rules of faultless liability, such as

---

social perception of our tort system."); Gary T. Schwartz, *The Ethics and the Economics of Tort Liability Insurance*, 75 CORNELL L. REV. 313, 327 (1990) [hereinafter Schwartz, *Liability Insurance*].

63. *See, e.g.*, Payton v. Abbott Labs, 437 N.E.2d at 176 (noting that tort liability has a retributive function).

64. ARIEL PORAT & ALEX STEIN, TORT LIABILITY UNDER UNCERTAINTY 104 (2001); ARTHUR RIPSTEIN, EQUALITY, RESPONSIBILITY, AND THE LAW 77 (1999) (stating that from a moral standpoint those who differ only by luck should fare equally well or badly); Kenneth Mann, *Punitive Civil Sanctions: The Middleground Between Criminal and Civil Law*, 101 YALE L.J. 1795, 1806 (1992) ("The criminal sanction will apply even if no individual interest has suffered direct injury. The paradigmatic civil sanction . . . applies to conduct that causes actual damage to an individual interest . . . ."); Richard W. Wright, *Substantive Corrective Justice*, 77 IOWA L. REV. 625, 668 (1992) [hereinafter Wright, *Corrective Justice*] (embracing Coleman's criticism of retributive theories). *See also* Schwartz, *Liability Insurance*, *supra* note 62, at 327, 328. Schwartz observes that "from the standpoint of retribution tort law behaves curiously" when it imposes liability only after injury occurs. *Id.* at 328. Schwartz points out that a perfectly responsive insurance scheme in which premiums reflect the exact level of risk that the insured's own conduct occasions may eliminate this oddity. *Id.* at 328 The effective burden borne by the insured becomes a function of his risk taking rather, than the fortuities of what accidents happen. *Id.* This kind of insurance, however, is unrealistic, as Schwartz himself recognizes. *Id.* at 320. Furthermore, responsive insurance, at least as Schwartz defines it, is responsive only to levels of risk and not to other determinants of the gravity of a wrongful conduct. *Id. Cf.* Christopher Schroeder, *Corrective Justice, Liability for Risks, and Tort Law*, 38 UCLA L. REV. 143, 152 (1990). Schroeder offers a "liability based on risk creation" scheme. However, as Schroeder himself admits, this theory does not square with retributive justice, because expected harm is not the only determinant of the gravity of one's wrong. *Id.* at 152-53. In any case, this theory is a proposal for reform; it does not intend to explain tort law as we know it.

65. Jeremy Waldron, *Moments of Carelessness and Massive Loss*, *in* PHILOSOPHICAL FOUNDATIONS OF TORT LAW 387, 401-05 (David G. Owen ed., 1995). Waldron argues that tort law exposes every wrongdoer "to a risk of liability exactly equivalent to the risk of loss that [she or he] imposed on [others]." *Id.* at 403.

66. First, as stated above, the magnitude of the risk is not the sole determinant of the gravity of one's wrong. Second, it is doubtful that the creation of a particular risk may be regarded as both the wrong and the sanction in a retributive model. I intend to elaborate on Waldron's theory in a future article.

strict and vicarious liability.[67] Strict liability is imposed for begetting harm. It is independent of moral wrongfulness unless one considers causation of harm to be wrongful *per se*. Vicarious liability is imposed on one person for the causation of harm by another. This is an even more peculiar doctrine from a retributive standpoint, because the sanction is borne by a person who neither committed a moral wrong nor caused any harm.

Moreover, even within the law of negligence, liability may be imposed for conduct that can hardly be considered culpable. For example, under the objective standard of reasonable conduct, even a person who lacks the subjective ability to understand or abide by the standard of reasonableness may be deemed negligent.[68] The negligence *per se* doctrine poses a similar problem.[69]

Third, even where wrongful conduct results in harm, the severity of the sanction imposed by tort law is determined by the fortuitous amount of the plaintiff's loss, which is usually a poor measure of the gravity of the defendant's wrong.[70] The extent of tort liability may be incompatible with the principles of cardinal and ordinal proportionality. A slight and absentminded deviation from the objective standard of care may result in serious injury and may therefore lead to extensive liability. This outcome violates the retributive principle of cardinal proportionality. A similar violation of this principle occurs where a considerable and knowing deviation from the standard of care results in a trivial injury.

Furthermore, a loss-based sanctioning system violates the principle of ordinal proportionality. Assume that A negligently injures B, and C negligently injures D, and that B's physical injury is identical to D's injury. If A and C committed indistinguishable wrongful acts, but B's earning capacity is higher than D's, then retributive justice would demand similar sanctions on A and C; whereas, tort law will impose a more severe monetary sanction on A. Yet if B

---

67. Schwartz, *Liability Insurance*, *supra* note 62, at 326-27; Steven D. Smith, *The Critics and The "Crisis": A Reassessment of Current Conceptions of Tort Law*, 72 CORNELL L. REV. 765, 776 (1987); Stephen D. Sugarman, *Doing Away with Tort Law*, 73 CAL. L. REV. 555, 610 (1985). One may refine one's definition of the role of retribution in tort law by distinguishing two branches of liability: fault liability which is rooted in retribution, and strict liability which is based on some other non-retributive goal. Cane, *supra* note 20, at 159-60, 165, 171. This refinement, however, does not invalidate the other criticisms of the "retributive mechanism" perception.

68. Cane, *supra* note 20, at 141-42; Smith, *supra* note 67, at 776; Sugarman, *supra* note 67, at 610; Wright, *Corrective Justice*, *supra* note 64, at 668.

69. Under this doctrine, an unexcused violation of a legislative act or an administrative regulation, may in itself constitute negligence. RESTATEMENT (SECOND) OF TORTS §§ 286, 288B (1965).

70. PORAT & STEIN, *supra* note 64, at 104; Cane, *supra* note 20, at 142; John G. Fleming, *The Collateral Source Rule and Loss Allocation in Tort Law*, 54 CAL. L. REV. 1478, 1483-84 (1966); David G. Owen, *Deterrence and Desert in Tort: A Comment*, 73 CAL. L. REV. 665, 668-69 (1985); Stephen R. Perry, *The Moral Foundations of Tort Law*, 77 IOWA L. REV. 449, 470 (1992); Schwartz, *Liability Insurance*, *supra* note 62, at 327; Smith, *supra* note 67, at 776-77; Sugarman, *supra* note 67, at 610; Wright, *Corrective Justice*, *supra* note 64, at 668.

and D have equal earning capacities, but A's conduct is more culpable than C's, then retributive justice mandates a stricter sanction on A, while tort law imposes similar sanctions on A and C.

Fourth, retribution insists on penalizing the actual wrongdoer. The burden must be borne by the one who deserves to bear it. Yet, even in instances where a wrongful act has been committed, the wrongdoer does not always bear the burden of tort liability. Very often, liability insurance removes the burden of liability from the actual wrongdoer.[71] In many cases, the wrongdoer's employer bears the cost.[72] Although higher insurance rates and job penalties may punish some wrongdoers to some extent, these unofficial sanctions are erratic and rarely correspond to the gravity of the conduct.[73] Ultimately, the wrongdoer does not get her due.

B. The Role of Retributive Justice from a Pluralistic Perspective

So far I have shown that retributive justice cannot be regarded as the principal rationale for tort liability. However, this does not mean that the notion of retributive justice is completely absent from tort jurisprudence. Next, I shall put forward a twofold thesis. My first assertion is that retributive justice plays a certain role in tort law. Imposition, or expansion, of tort liability is a penalty on the defendant, and exclusion, or limitation, of liability is a penalty on the plaintiff. If these penalties were at times insupportably unfair from a retributive standpoint, people would probably feel that there was something wrong with the system that imposed them. Given that the idea of retribution is deeply rooted in the moral intuition of every person, including judges and jurors, one would be very surprised if there were no sign of it in tort law.

My second assertion is that the use of retributive rhetoric, or logic, in the common law of torts is reserved for fairly limited circumstances that can be subsumed under two paradigms. This is also understandable. Tort law is concerned with bilateral settings in which one person has caused harm to another and is asked to repair that harm. It is, therefore, closely linked to the notion of corrective justice.[74] A systematic attempt to scrutinize tort sanctions through the prism of retributive justice would undoubtedly undermine tort law's corrective structure. The need to preserve this fundamental structure reduces the importance that can be attached to retributive concerns in tort adjudication. Ergo, the expected role of retributive justice in tort law must be limited.

As will be shown below, retributive justice is one of the leading explanations, whether explicit or implicit, for certain components of tort

---

71. Smith, *supra* note 67, at 795; Sugarman, *supra* note 67, at 573-81, 609.

72. Sugarman, *supra* note 67, at 609.

73. *Id.*

74. *See, e.g.*, Izhak Englard, *The Idea of Complementarity as a Philosophical Basis for Pluralism in Tort Law*, *in* PHILOSOPHICAL FOUNDATIONS OF TORT LAW 183, 194 (David G. Owen ed., 1995) ("In the light of the initial bilateral setting of tort adjudication, special weight should be given to the notion of corrective justice"); Schwartz, *Mixed Theories*, *supra* note 2, at 1816.

doctrine. This does not necessarily mean that retributive justice receives its proper weight in tort law. Some may argue that it is given too much weight, others that its use is overly restrained. However, this article focuses on the actual utilization of retributive concerns in tort jurisprudence, and not on their precise normatively defensible role. The latter issue must await further research.

As tort adjudication links two persons, and given that there are two possible deviations from the principle of retributive justice (an excessively severe or a too lenient sanction), this principle may be utilized in four ways, at least theoretically. There may be a *contraction* of liability where the severity of the *defendant's* conduct justifies a more lenient sanction than that prescribed by the notion of corrective justice; *contraction*, or even exclusion, of liability where the severity of the *plaintiff's* conduct entails such a response; *expansion* of liability where the gravity of the *defendant's* wrong requires a more severe sanction than that prescribed by the notion of corrective justice; and imposition or *expansion* of liability where corrective justice would leave an unduly onerous burden on the shoulders of the *plaintiff*.

Tort law is in fact responsive to retributive concerns within two conceptual paradigms:

- *Prevention of abominable disproportion.* Tort law can hardly ever impose a sanction that accurately fits the wrong. Disproportion between the gravity of the wrong and the severity of the sanction is an inevitable consequence of the primacy of corrective justice. Yet, tort law does not seem to tolerate especially alarming disparities between sanctions and wrongs. It endeavors to prevent the imposition of sanctions that are *abominably* too severe or too lenient, in light of the respective wrongs. The application of corrective justice is thereby limited when it leads to extreme injustice from a retributive perspective.
- *Preservation of criminal justice.* Tort law sometimes appears to function as a servant of criminal law.[75] Several tort doctrines may be explained as an attempt to preserve and vindicate criminal justice, as prescribed by the principles and rules of criminal law and procedure. The principal justifications for criminal punishment are currently understood to be retribution and deterrence.[76] Consequently, when tort law vindicates criminal justice, it may be said to further these twin goals.

IV. THE ROLE OF RETRIBUTIVE JUSTICE IN LIMITING LIABILITY

A. Protecting the Injurer from an Excessive Sanction

1. The Main Idea

The concept of retributive justice cannot explain tort liability; tort law is basically a manifestation of corrective justice. It surely penalizes the

75. *See supra* Subsection IV.B.1.c and Section IV.B.2.

76. *See* sources cited *supra* note 2.

wrongdoer but it also compensates the immediate victim of the wrong. One is interlinked with the other. Trying to punish the wrongdoer according to the principle of retributive justice may result in under-compensation or over-compensation in nearly all cases. Nonetheless, retributive justice is applied by the courts in extreme circumstances, where it is thought that imposing liability according to the principle of corrective justice may result in an *abominable* disproportion between the burden of civil liability and the severity of the defendant's conduct. This utilization of retributive justice has two aspects. Courts find it unfair to impose an extremely onerous burden on a person whose only mischief was a slight absentminded deviation from an objective standard of care.[77] At the same time, they find it iniquitous to let the doer of an extremely shocking and intentional act pay only for the actual loss caused, especially where, fortuitously, the loss is not serious.

The first component of the prevention-of-abominable-disproportion argument seems to be one of the most plausible justifications for two common law rules that exclude, or limit, liability for relational economic loss and relational emotional harm. Relational economic loss may be defined as financial loss consequent upon negligent infliction of harm to the person or property of a third party, or to an ownerless tangible resource. The same definition applies, *mutatis mutandis*, to relational emotional injury. These two topics shall be discussed consecutively.

2. Relational Economic Loss

Liability for relational economic loss is excluded in most common law jurisdictions. In another article, I have elaborated on the history and rigid application of the "exclusionary rule."[78] Some of the most frequently cited justifications for this rule rest on the fear of open-ended liability. In the seminal case of *Ultramares Corp. v. Touche*,[79] which is not a relational loss case, the late Justice Cardozo observed that allowing claims for pure economic loss may expose the wrongdoer to "liability in an indeterminate amount for an indeterminate time to an indeterminate class."[80] Nevertheless, the same

77. *See* Honoré, *supra* note 20, at 89 ("[T]he compensation payable may be disproportionate to what is often a minor fault. To avoid this disproportion, the retributive principle seems to require that defendants should not be exposed to disproportionately heavy losses"). Honoré posits that insurance may be a needed solution for the problem of disproportionate burden. *Id.* at 89-90. However, insurance is not always obtainable, especially where the risk is indeterminate.

78. *See generally* Ronen Perry, *Relational Economic Loss: An Integrated Economic Justification for the Exclusionary Rule*, 56 RUTGERS L. REV. 711 (2004).

79. 174 N.E. 441 (N.Y. 1931).

80. *Id.* at 444. This case was a negligent misrepresentation case. *Id.* at 442. Defendants had been employed by a third-party to prepare and certify a balance sheet exhibiting the condition of third-party's business. *Id.* Defendants were aware that a third-party would use its certificate of audit to obtain credit for the operation of its business. *Id.* Although capital and surplus were certified to be intact, "in reality, both had been wiped out and the business was insolvent." *Id.* On the faith of defendants' certificate, plaintiff made several loans to defendant.

rationale has been invoked in numerous relational loss cases as the principal reason for exclusion of liability.[81] Two questions must be answered in this respect. First, is the fear of open-ended liability reasonable? Second, why should the likelihood of indeterminate liability result in exclusion of liability for relational losses?

The answer to the *first* question seems self-evident. A negligent infliction of injury to one person may result in economic loss to her relatives, customers, creditors, suppliers, employers, partners, and so on;[82] the economic loss of any entity may economically affect others without any foreseeable end. Similarly, injuring a factory may cause economic loss to its suppliers of raw materials, distributors, and consumers.[83] If the halted factory was manufacturing components for products assembled in another factory, the latter may also suffer economic loss.[84] Employers of the halted factories may lose their income, at least temporarily;[85] further, owners of shops and restaurants where those workers or their dependants customarily shop and dine may lose profits, and so on.[86]

Relational losses may spread serially or in parallel. For instance, when an electricity cable is damaged, resulting in total blackout in an industrial area, the economic losses of the halted plants are parallel; they all ensue from a single injury, the damage caused to the property of the electricity company, not from each other. When a railway bridge owned by the state is harmed, the lost profits of the private railway company using the bridge, and the additional outlays incurred by the owner of the cargo that cannot reach its destination in the usual way, are serially linked. The loss of the cargo owner results from the inability of the railway company to function, which was caused by the damage to the bridge.

In most cases, however, relational losses grow in complex patterns. Any

---

Id. at 443. Plaintiff brought an action against the defendants to recover the loss it suffered in reliance upon the audit. *Id.*

81. *See* Barber Lines A/S v. M/V Donau Maru, 764 F.2d 50, 54 (1st Cir. 1985); *In re* Waterstand Marine, Ltd., No. 87-1516, 1988 U.S. Dist. LEXIS 3242, at *12-13 (E.D. Pa. 1988); Pruitt v. Allied Chem. Corp., 523 F. Supp. 975, 979-80 (E.D. Va. 1981); Byrd v. English, 43 S.E. 419, 420 (Ga. 1903).

82. *See*, Champion Well Serv., Inc. v. NL Indus., 769 P.2d 382, 385 (Wyo. 1989) (holding that an employer cannot recover economic losses consequent upon a negligent infliction of harm to its key employee).

83. *See, e.g.*, PPG Indus., Inc. v. Bean Dredging, 447 So. 2d 1058, 1061-62 (La. 1984) (denying recovery by a customer of the gas company when a gas pipeline was negligently injured by the defendant, and the customer was required to obtain gas from another source during the repairs at an increased cost); *see also* J.A. Smillie, *Negligence and Economic Loss*, 32 U. TORONTO L.J. 231, 241 (1982) (illustrating that interruption of production in one factory may cause economic loss to those who supply it with raw materials, those who distribute the products and those who purchase its products).

84. Smillie, *supra* note 83, at 241.

85. *See, e.g.*, Stevenson v. East Ohio Gas, Co., 73 N.E.2d 200, 203-04 (Ohio Ct. App. 1946).

86. Smillie, *supra* note 83, at 241.

relational loss, whether or not it has parallels, may generate further losses, which are parallel but serially linked to their antecedent. Each of the subsequent losses may, itself, become the source of further losses. For example, marine oil pollution may cause economic loss to fishermen, oystermen, crabbers, and their like.[87] These losses are parallel. The inability of a certain fisherman to work may cause economic loss to seafood restaurants or retailers that generally buy his catch.[88] Their losses are parallel and serially linked to the fisherman's loss. The inability of a seafood merchant to buy from fishermen may also cause financial loss to fish canneries, fish processing factories, and more.[89] Theoretically, such proliferation of economic losses is boundless, so the number of potential relational victims is vast and indeterminate. This phenomenon has rightly been termed "the ripple effects,"[90] "the domino effect,"[91] or the "chain reaction."[92]

The potential number of victims may, in itself, have some normative significance, but its relevance largely depends on the rough correlation between the number of valid claims and the extent of tort liability.[93] The larger the number of valid claims the more extensive the liability; if the potential number of victims is indeed large and uncertain, then potential liability is also large and uncertain.

Now, the second question: *why* is the likelihood of open-ended liability deemed normatively relevant? Several answers are possible;[94] however, one is of special interest to this article. It is very frequently said by Commonwealth jurists that allowing recovery for relational losses may give rise to an

---

87. *See, e.g.*, Union Oil Co. v. Oppen, 501 F.2d 558 (9th Cir. 1974).

88. *Cf. In Re* Ballard Shipping Co. v. Beach Shellfish, 32 F.3d 623, 625 (1st Cir. 1994) (denying claims by seafood dealers, tackle shop operators, restaurant owners and employees, a scuba equipment and canoe rental shop, and a variety of other shoreline businesses for economic loss arising from an oil spill).

89. *Cf. In Re* Exxon Valdez, 1994 U.S. Dist. LEXIS 20555, at *7 (D. Alaska 1994) (denying claims by seafood wholesaler, processor, cannery employee, and tenderer for economic loss resulting from an oil spill).

90. Can. Nat'l Ry. Co. v. Norsk Pac. S.S. Co., [1992] 91 D.L.R.(4th) 289, 302; Jane Stapleton, *Duty of Care and Economic Loss: A Wider Agenda*, 107 L.Q. REV. 249, 255 (1991).

91. David R. Owen, *Recovery for Economic Loss Under U.S. Maritime Law: Sixty Years Under Robins Dry Dock*, 18 J. MAR. L. & COM. 157, 163 (1987); Carl F. Stychin, *"Principled Flexibility": An Analysis of Relational Economic Loss in Negligence*, 25 ANGLO-AM. L. REV. 318, 319 (1996).

92. W. PAGE KEETON, ET AL., PROSSER & KEETON ON THE LAW OF TORTS 1001 (5th ed. 1984).

93. *See, e.g.*, Stevenson v. East Ohio Gas, Co., 73 N.E.2d at 203 ("[T]o permit recovery of damages . . . would open the door to a mass of litigation which might very well overwhelm the courts . . . ."); Perry, *supra* note 78, at 761-63 (discussing the benefits of primary loss-spreading to society).

94. For example, it may be said that as the extent of tort liability grows the marginal deterrent effect diminishes, either because no further precautions are available or because the total amount of the claims exceeds the upper limit of the injurer's liability (natural or legal). Louisiana *ex rel.* Guste v. M/V Testbank, 752 F.2d 1019, 1029 (5th Cir. 1985).

abominable disproportion between the severity of the sanction and the gravity of the wrong, given that liability may be unlimited.[95] An insignificant and perhaps absentminded deviation from the objective standard of care cannot justify the imposition of such an onerous penalty. As stated by Justice Gibbs in the well-known Australian case of *Caltex Oil (Austl.) Pty. Ltd. v. The Dredge "Willemstad"*[96]

> [I]f through the momentary inattention of an officer, a ship collided with a bridge, and as a result a large suburban area, which included shops and factories, was deprived of its main means of access to a city, great loss might be suffered by tens of thousands of persons, but to require the wrongdoer to compensate all those who had suffered pecuniary loss would impose upon him a burden out of all proportion to his wrong.[97]

This type of reasoning is also employed by American jurists. For example, in *Phoenix Prof'l Hockey Club, Inc. v. Hirmer*,[98] the Supreme Court of Arizona observed that the imposition of liability for relational losses "could impose a severe penalty on one guilty of mere negligence."[99] Similarly, the Superior Court of Pennsylvania concluded in *Aikens v. Baltimore & Ohio R.R. Co.*,[100] that imposition of such liability "would create a disproportion between the large amount of damages that might be recovered and the extent of the defendant's fault."[101] The Second Restatement of Torts adopted the exclusionary rule,[102]

---

95. *Can. Nat'l Ry. Co.*, 91 D.L.R. (4th) at 365-66; Leigh & Sillavan, Ltd. v. Aliakmon Shipping Co., (1986) 2 All E.R. 145, 154 (H.L.); Burgess v. Florence Nightingale Hosp. for Gentlewomen, (1955) 1 Q.B. 349, 355-56; Nacap Ltd. v. Moffat Plant Ltd. 1987 S.L.T. 221, 222; R.P. BALKIN & J.L.R. DAVIS, LAW OF TORTS 422 (1991); Robert Hayes, *The Duty of Care and Liability for Purely Economic Loss*, 12 MELB. U. L. REV. 79, 82 (1979); D. Marshall, *Liability for Pure Economic Loss Negligently Caused—French and English Law Compared*, 24 INT'L & COMP. L.Q. 748, 750 (1975); Smillie, *supra* note 83, at 231.

96. (1976) 136 C.L.R. 529 (Austl.).

97. *Id.* at 551; *see id.* at 562-63 (Stephen J., concurring), 591 (Mason J., concurring). For further analysis of the *Caltex* decision *see* Perry, *supra* note 78, at 724.

98. 502 P.2d 164, (Ariz. 1972) (dismissing an action by an employer to recover out-of-pocket expenses in hiring a replacement for an employee injured as a result of defendant's negligence).

99. *Id.* at 165.

100. 501 A.2d 277 (Pa. Super. Ct. 1985).

101. *Id.* at 279. *See* Gen. Foods Corp. v. United States, 448 F.Supp. 111, 112-13 (D. Md. 1978); Stromer v. Yuba City, 37 Cal. Rptr. 240, 243-44 (Cal. App. 1964); People Express Airlines, Inc. v. Consol. Rail Corp., 495 A.2d 107, 110 (N.J. 1985); Rickards v. Sun Oil Co., 41 A.2d 267, 269 (N.J. 1945). *But see* Kinsman Transit Co. v. City of Buffalo, 388 F.2d 821 (2nd Cir. 1968) (observing the prominence of the disproportion argument, but declining to adopt it).

102. RESTATEMENT (SECOND) OF TORTS § 766C (1979). This section excludes liability for negligent interference with contract or prospective contractual relation. *Id.* It does not deal with "pure" economic expectations that are not based on existing or expected contracts. *Id.* at Comment a. However, it seems that if tort law does not protect contractual expectations from negligent interference by third parties, then it does not protect pure expectations. Note that the drafts of the third Restatement of Torts do not currently consider liability for non-physical harm,

and specifically embraced the likelihood of abominable disproportion as one of its principal justifications.[103] The same reasoning has been cited by numerous tort scholars.[104]

The prevention-of-abominable-disproportion argument is obviously inapplicable where the extent of potential liability is limited *ex lege*. For example, the liability of ship owners for losses caused by collision has been limited, by statute, for centuries in most Western jurisdictions.[105] Currently, there are two established methods for such limitation. In the United States, for example, the owner's liability cannot exceed the amount or value of her interest in the vessel, and its freight then pending.[106] In Australia, Canada, and England, on the other hand, the maximal liability of the owner is proportionate to the tonnage of the vessel.[107] In the twentieth century, liability of airlines was limited in a similar manner.[108] Whenever potential liability is limited, liability

---

although § 766C is occasionally mentioned. See RESTATEMENT (THIRD) OF TORTS: LIABILITY FOR PHYSICAL HARM § 6 cmt. d, and § 29 cmt. a (Proposed Final Draft No. 1 2005).

103. RESTATEMENT (SECOND) OF TORTS § 766C cmt. a (1979) ("[C]ourts apparently have been influenced by . . . the probable disproportion between the large damages that might be recovered and the extent of the defendant's fault . . . .").

104. Henry D. Gabriel, *Testbank: The Fifth Circuit Reaffirms The Bright Line Rule of Robins Dry Dock And Fails to Devise a Test to Allow Recovery for Pure Economic Damages*, 31 LOY. L. REV. 265, 266 (1985); Ann O'Brien, *Limited Recovery Rule as a Dam: Preventing a Flood of Litigation for Negligent Infliction of Pure Economic Loss*, 31 ARIZ. L. REV. 959, 967 (1989); John G. Rich, *Negligent Interference with Prospective Economic Advantage - J'Aire Corp. v. Gregory*, 1980 UTAH L. REV. 431, 434; Kelly M. Hnatt, Note, *Purely Economic Loss: A Standard for Recovery*, 73 IOWA L. REV. 1181, 1183, 1193 (1988); Comment, *Foreseeability of Third Party Economic Injuries - A Problem in Analysis*, 20 U. CHI. L. REV. 283, 296, 298 (1953); Note, *Negligent Interference with Contract: Knowledge as a Standard for Recovery*, 63 VA. L. REV. 813, 817 (1977); Recent Cases, *Torts - Interference with Business or Occupation - Commercial Fishermen Can Recover Profits Lost as a Result of Negligently Caused Oil Spill*, 88 HARV. L. REV. 444, 448 (1974); *see also* Robert L. Rabin, *Tort Recovery for Negligently Inflicted Economic Loss: A Reassessment*, 37 STAN. L. REV. 1513, 1534, 1538 (1985) (asserting that the abhorrence to disproportionate penalties for wrongful behavior is the most plausible explanation for the judicial reluctance to allow recovery for pure economic loss).

105. G. Todd Stanley, *Economic Loss in Maritime Law: On Course for Reevaluating the Exclusionary Rule*, 53 U. TORONTO FAC. L. REV. 312, 340-42 (1995).

106. 46 U.S.C.. app. § 183(a) (2000).

> The liability of the owner of any vessel . . . for any embezzlement, loss, or destruction by any person of any property, goods, or merchandise shipped or put on board of such vessel, or for any loss, damage, or injury by collision . . . shall not . . . exceed the amount or value of the interest of such owner in such vessel, and her freight then pending.

Id.; *see also* THOMAS J. SCHOENBAUM, ADMIRALTY AND MARITIME LAW 808-33 (3rd ed. 2001).

107. Limitation of Liability for Maritime Claims Act, 1989, art. 6 (Austl.); Marine Liability Act, S.C. 2001, c. 6, §§ 24-34 (Can.); Merchant Shipping Act, 1995, c. 21, § 191 (Eng.). These statutes embrace the International Convention on Limitation of Liability for Maritime Claims, Nov. 19, 1976, 16 I.L.M. 606, 608.

108. Civil Aviation (Carriers' Liability) Act, 1959, § 11A (Austl.) (limiting carrier liability to 260,000 SDRs as a general rule); Carriage by Air Act, R.S.C. ch. C-26, art. 22 sched. 1 (1985) (Can.) (limiting carrier liability to the sum of 125,000 francs for each passenger);

insurance is readily available, and abominable disproportion is highly unlikely. In these cases, however, the exclusionary rule may be supported by other legitimate considerations.[109]

3. Relational Emotional Harm

Relational emotional harm is an emotional distress consequent upon a negligent infliction of harm to the person, or property, of another. It may result from: witnessing the occurrence in which a third party, or some property, is injured or imperiled; learning about the ensuing injury after its occurrence; sympathizing with the mental anguish of another relational victim. According to the traditional rule of the common law, relational emotional harm was irrecoverable. At the beginning of the twentieth century, liability for negligent infliction of emotional harm was recognized in the United States only in so far as that harm culminated in illness or bodily impact.[110] Pure emotional (intangible) harm was irrecoverable.[111] Moreover, even when emotional distress culminated in physical manifestation, the victim could not recover if it was occasioned by fear of injury to property or to the person of another.[112]

Most, if not all, jurisdictions eventually recognized at least a limited exception to the exclusionary rule. The most restrictive approach allows a person to recover for emotional harm that she suffered from witnessing another person being injured in a certain incident only if she was physically injured in the same incident. This has been called the *physical-impact* doctrine. Today, this rule prevails in few jurisdictions.[113] A more generous approach, once conceived as the majority view in the United States, enables a person to recover for relational emotional harm if she was among those who were physically

---

Carriage by Air Act, 1961, 9 & 10 Eliz. 2, c. 27, § 4 sched. 1 (Eng.) (limiting carrier liability by aggregate number of passengers).

109. For example, if there is a limited pool that all valid claims need to share, courts may deny recovery for relational losses in order to guarantee full recovery for injuries to physical interests which are deemed more worthy of legal protection.

110. RESTATEMENT (SECOND) OF TORTS § 313(1) (1965) ("If the actor unintentionally causes emotional distress to another, he is subject to liability to the other for resulting illness or bodily harm if [certain conditions are met] . . . .").

111. RESTATEMENT (SECOND) OF TORTS § 436A (1965) ("If the actor's conduct is negligent as creating an unreasonable risk of causing either bodily harm or emotional disturbance to another, and it results in such emotional disturbance alone, without bodily harm or other compensable damage, the actor is not liable for such emotional disturbance.").

112. S. Ry. Co. v. Jackson, 91 S.E. 280 (Ga. 1916) (finding no recovery for fright induced by the apprehension for another person's safety); Cleveland, C., C. & St. L. Ry. Co. v. Stewart, 56 N.E. 917, 922 (Ind. App. 1900) (stating that mere fright was not enough to rise to the level to incur damages); Mahoney v. Dankwart, 79 N.W. 134, 136 (Iowa 1899) (finding no recovery where physical injury was caused by fright).

113. E.g., Lee v. State Farm Mut. Ins. Co., 533 S.E.2d 82, 85 (Ga. 2000); Kraszewski v. Baptist Med. Ctr. of Okla., Inc., 916 P.2d 241, 248 (Okla. 1996); Sherwood v. Or. Dept. of Transp., 11 P.3d 664, 671 (Or. Ct. App. 2000).

endangered by the conduct that caused the physical injury to the third party.[114] This is perceived as the *zone-of-danger* doctrine, and it still holds in many jurisdictions.[115] In its narrow form, this doctrine applies only where the emotional harm culminates in physical manifestation.[116] In its broader form, the zone of danger doctrine applies to any *serious* emotional harm.[117] From a logical perspective, the zone-of-danger exception encompasses (and is therefore broader than) the physical-impact exception.[118]

In many common law jurisdictions liability for relational emotional harm was expanded even further. *Dillon v. Legg*[119] was the turning point. The Supreme Court of California overruled its earlier decisions, holding that liability for relational emotional harm should depend on its foreseeability.[120] In determining whether the harm was reasonably foreseeable, courts should consider: (1) "[w]hether the plaintiff was located near the scene of the accident or a distance away from it"; (2) whether or not the plaintiff's shock resulted from the "sensory and contemporaneous observance of the accident"; (3) whether or not the plaintiff and the victim were "closely related."[121] The court stated that "the evaluation of these factors [would] indicate the *degree* of the defendant's foreseeability" in each case.[122] The court confined its ruling to cases in which the plaintiff's fear was manifested in physical consequences.[123] This requirement was eventually replaced by the need to show that the plaintiff's emotional distress was serious.[124]

More than half of the states have adopted *Dillon's* holding that recovery for

---

114. *See* RESTATEMENT (SECOND) OF TORTS §§ 313(2), 436(3) (1965).

115. These jurisdictions include Alabama, Arizona, District of Columbia, Illinois, Minnesota, Missouri, New York, North Dakota, South Dakota, Vermont, Utah, and Washington. AALAR, Ltd. v. Francis, 716 So. 2d 1141, 1147 (Ala. 1998); Keck v. Jackson, 593 P.2d 668, 669 (Ariz. 1979); Johnson v. District of Columbia, 728 A.2d 70, 77 (D.C. 1999); Rickey v. Chi. Transit Auth., 457 N.E.2d 1, 5 (Ill. 1983); K.A.C. v. Benson, 527 N.W.2d 553, 557, 558 (Minn. 1995); Asaro v. Cardinal Glennon Mem'l Hosp., 799 S.W.2d 595, 599-600 (Mo. 1990); Bovsun v. Sanperi, 461 N.E.2d 843, 848 (N.Y. 1984); Muchow v. Lindblad, 435 N.W.2d 918, 921 (N.D. 1989); Nielson v. AT&T Corp., 597 N.W.2d 434, 442 (S.D. 1999); Leo v. Hillman, 665 A.2d 572, 577 (Vt. 1995); Hansen v. Sea Ray Boats, Inc., 830 P.2d 236, 239 (Utah 1992); Kloepfel v. Bokor, 66 P.3d 630, 634-35 (Wash. 2003); *see also* Benjamin C. Zipursky, *Rights, Wrongs, and Recourse in the Law of Torts*, 51 VAND. L. REV. 1, 29 (1998) ("[T]he notorious "zone of danger' rule in bystander cases [has been] adopted by many courts including the United States Supreme Court.").

116. *Keck*, 593 P.2d at 669-70; *Rickey*, 457 N.E.2d at 5; *K.A.C.*, 527 N.W.2d at 557; *Muchow*, 435 N.W.2d at 922.

117. *Bovsun*, 461 N.E.2d at 849-50.

118. Any person who sustains a physical injury was obviously within the zone of physical danger.

119. 441 P.2d 912 (Cal. 1968).

120. *Id.* at 919-20.

121. *Id.* at 920.

122. *Id.*

123. *Id.*

124. Molien v. Kaiser Found. Hosps., 616 P.2d 813, 821 (Cal. 1980); Hedlund v. Superior Court, 669 P.2d 41, 47, 706 n.8 (Cal. 1983).

emotional distress suffered by bystanders is not limited to persons who suffer physical impact in the same incident or are in the zone of danger.[125] Among these states, only a few allow recovery on the basis of seriousness and foreseeability (with reference to the *Dillon* guidelines).[126] All other states hold that foreseeability does not suffice, and that other specific requirements must also be met. In many of them, including California, the *Dillon* guidelines have been practically converted, sometimes with variations or additions, into substantive limitations on recovery.[127] For example, it is generally necessary that the plaintiff have a close relationship with the injured victim.[128] Additionally, in most states, recovery is allowed only where the harm culminates in physical manifestation,[129] or is at least serious or severe.[130]

In summation, courts have always been reluctant to impose liability for relational emotional harm. In the past, courts excluded liability for these harms

---

125. Dale Joseph Gilsinger, Annotation, *Recovery Under State Law for Negligent Infliction of Emotional Distress Under Rule of Dillon v. Legg, 68 Cal. 2d 728, 69 Cal. Rptr. 72, 441 P.2d 912 (1968), or Refinements Thereof*, 96 A.L.R. 5th 107 (2002); Consol. Rail Corp. v. Gottshall, 512 U.S. 532, 549 (1994).

126. Montana, New Hampshire and North Carolina. Sacco v. High Country Indep. Press, 896 P.2d 411, 425 (Mont. 1995) (ruling that serious or severe harm which is reasonably foreseeable is recoverable); Graves v. Estabrook, 818 A.2d 1255, 1258 (N.H. 2003); Gardner v. Gardner, 435 S.E.2d 324, 327 (N.C. 1993).

127. Richard S. Miller, *The Scope of Liability for Negligent Infliction of Emotional Distress: Making "the Punishment Fit the Crime,"* 1 U. HAW. L. REV. 1, 5 (1979). For a general statement of the development of the zone of danger doctrine see *Consol. Rail Corp. v. Gottshall*, 512 U.S. 532, 549 (1994). With regard to California see *Thing v. La Chusa*, 771 P.2d 814, 817-25 (Cal. 1989).

128. *See* Dale Joseph Gilsinger, Annotation, *Relationship between Victim and Plaintiff-Witness as Affecting Right to Recover under State Law for Negligent Infliction of Emotional Distress Due to Witnessing Injury to Another Where Bystander Plaintiff Is Not Member of Victim's Immediate Family*, 98 A.L.R.5th 609 (2002) (discussing cases involving different levels of relationship). *But see* Lourcey v. Estate of Scarlett, 146 S.W.3d 48, 53-54 (Tenn. 2004) (finding that plaintiff is not required to establish a close relationship to the injured party in order to have a cause of action).

129. *E.g.*, Zell v. Meek, 665 So. 2d 1048, 1054 (Fla. 1995); Taylor v. Kurapati, 600 N.W.2d 670, 693 (Mich. Ct. App. 1999); Folz v. State, 797 P.2d 246, 260 (N.M. 1990); Swerdlick v. Koch, 721 A.2d 849, 864 (R.I. 1998); Stewart v. State Farm Mut. Auto. Ins. Co., 533 S.E.2d 597 (S.C. Ct. App. 2000).

130. *E.g.*, M.A. v. United States, 951 P.2d 851, 856 (Alaska 1998); Clohessy v. Bachelor, 675 A.2d 852, 864 (Conn. 1996); Groves v. Taylor, 729 N.E.2d 569, 573 (Ind. 2000); Trahan v. McManus, 728 So. 2d 1273, 1279 (La. 1999); Michaud v. Great N. Nekoosa Corp., 715 A.2d 955, 959 (Me. 1998); Vosburg v. Cenex-Land O'Lakes Agronomy Co., 513 N.W.2d 870, 873 (Neb. 1994); *Lourcey*, 146 S.W.3d at 52; Hegel v. McMahon, 960 P.2d 424 (Wash. 1998); Jones v. Sanger, 512 S.E.2d 590, 595 (W. Va. 1998); Bowen v. Lumbermens Mutual Casualty Co., 517 N.W.2d 432, 434 (Wisc. 1994). A few state courts do not require that the plaintiff's harm physically manifest itself or be serious or severe. *See, e.g.*, Grotts v. Zahner, 989 P.2d 415, 416130. (Nev. 1999) (setting forth elements without including severity of the harm); Contreras v. Carbon County Sch. Dist. #1, 843 P.2d 589, 593 (Wyo. 1992) (omitting severity of harm from list of necessary elements).

altogether. Some jurisdictions still adhere to the exclusionary rule, subject to a limited exception, such as the physical-impact exception or the zone-of-danger exception. Others currently resolve cases of relational emotional harm with concrete limiting formulas. Only a few jurisdictions allow recovery on the basis of severity and foreseeability, but seem hesitant to conclude that emotional harm is indeed foreseeable unless the circumstances are such that liability could have been allowed in other jurisdictions as well.[131]

Relational emotional harms, just like relational economic losses, tend to ripple. Whenever a person is negligently injured, many others may suffer emotional distress. The first cycle of relational victims includes those who witnessed the accident or its consequences, be they relatives, friends, neighbors, or casual bystanders. The second cycle includes those who learned about the injury from another source, who was or was not a bystander. The third cycle includes those who sympathize with the mental anguish of other relational victims. A single physical injury may bring about numerous relational emotional afflictions. Once again, although the potential number of victims may, in itself, have some normative significance, its relevance greatly depends on the correlation between the number of valid claims and the extent of liability. If the potential number of victims is large and uncertain, the scope of potential liability is also large and uncertain.

Consequently, the rhetoric of the courts in cases of relational emotional harm has been somewhat similar to that employed in cases of relational economic loss. Courts have frequently observed that allowing recovery for relational emotional harm may result in an unfair sanction, out of proportion to the gravity of the wrong committed. This line of reasoning was and is still being used by the proponents of all methods of limitation. For example, in its well-known decision in *Tobin v. Grossman*[132] the Court of Appeals of New York concluded that relational emotional harm must be irrecoverable,[133] stating that "[e]very injury has ramifying consequences, like the ripplings of the waters, without end,"[134] and that allowing recovery may give rise to "unduly burdensome liability."[135] Over twenty years later, the Court of Appeals opened an avenue to bystander recovery, but due to the weighty influence of the *Tobin* reasoning, expansion was very limited.[136] Likewise, the Supreme Court of Georgia very recently adopted the physical-impact doctrine.[137] It apparently

131. See, for example, *Gardner*, 435 S.E.2d at 328, where the Supreme Court of North Carolina observed that a plaintiff's absence from the scene at the time of the tort, although not decisive, militates against the foreseeability of her resulting emotional distress.

132. 249 N.E.2d 419 (N.Y. 1969).

133. *Id.* at 419-20.

134. *Id.* at 424.

135. *Id.* at 423.

136. Bovsun v. Sanperi, 461 N.E.2d 843 (N.Y. 1984).

137. Lee v. State Farm Mut. Ins. Co., 533 S.E.2d 82, 85 (Ga. 2000). According to the traditional approach in Georgia, mental suffering was recoverable only where it resulted from the physical injury sustained by the physical impact. *Id.* at 84. The court in *Lee* stated that

[w]hen . . . a parent and child sustain a direct physical impact and physical injuries

believed that "to avoid. . . liability out of all proportion to the degree of a defendant's negligence, . . . the right to recover for negligently caused emotional distress must be limited [in this manner]."[138]

The history of the zone-of-danger doctrine also revolves around the fear of abominable disproportion. In *Waube v. Warrington*,[139] for example, a woman "was looking out the window of her house watching her child cross the highway, and witnessed the negligent killing of the child by the defendant"; she died soon after because of emotional upset.[140] The Supreme Court of Wisconsin stated that a defendant's duty cannot "be extended to any recovery for physical injuries sustained by one out of the range of ordinary physical peril as a result of the shock of witnessing another's danger."[141] The court opined that allowing recovery for relational emotional harm would impose liability that is "wholly out of proportion to the culpability of the negligent tort-feasor."[142] This paragraph was cited with approval by the Supreme Court of California as one of the justifications for denying recovery outside the zone-of-danger exception in *Amaya v. Home Ice, Fuel & Supply Co.*,[143] only five years prior to *Dillon*.

Both Wisconsin and California have departed from the restrictive approach employed in *Waube* and *Amaya*,[144] but the abominable disproportion reasoning has been applied by many courts that still adhere to this time-honored methodology. For example, although the Court of Appeals of New York eventually lifted the absolute bar to bystander recovery, it adopted a rather narrow version of the zone-of-danger doctrine[145] under the influential caveat laid down in *Tobin*.[146] Similarly, the Court of Appeals of Arizona recently upheld a constricted version of the zone-of-danger exception to the traditional rule of no recovery,[147] stating that "[a] dominant concern has been the

---

through the negligence of another, and the child dies as the result of such negligence, the parent may attempt to recover for serious emotional distress from witnessing the child's suffering and death without regard to whether the emotional trauma arises out of the physical injury to the parent.

Id. at 86-87.

138. *Id.* at 86 n.7.

139. 258 N.W. 497 (Wis. 1935).

140. *Id.* at 497.

141. *Id.* at 501.

142. *Id.*

143. 379 P.2d 513, 524-25 (Cal. 1963).

144. Thing v. La Chusa, 771 P.2d 814, 829; Bowen v. Lumbermens Mutual Casualty Co., 517 N.W.2d 432, 434-35, 436, 442.

145. Trombetta v. Conkling, 626 N.E.2d 653, 656 (N.Y. 1993) (finding that victim's aunt is not a member of immediate family); Bovsun v. Sanperi, 461 N.E.2d 843, 848 (allowing recovery only if the plaintiff was (1) within the zone of danger and (2) a member of the direct victim's *immediate* family).

146. *Trombetta*, 626 N.E.2d at 656; *Bovsun*, 461 N.E.2d at 847; Tobin v. Grossman, 249 N.E.2d 419, 424.

147. Hislop v. Salt River Project Agr. Imp. & Power Dist., 5 P.3d 267, 272 (Ariz. Ct. App. 2000).

perceived need to maintain a proportionate economic relationship between liability and culpability, the failure to do which underlies much of the criticism of the foreseeability test."[148] Finally, the Supreme Court of the United States applied similar reasoning in *Consolidated Rail Corp. v. Gottshall,*[149] where it held that liability for negligent infliction of emotional distress under the Federal Employers' Liability Act (FELA)[150] was limited to plaintiffs who were within the zone-of-danger.[151]

Even in its radical decision in *Dillon*, the Supreme Court of California explained that there was a need to "limit the otherwise potentially infinite liability which would follow every negligent act."[152] The court believed, however, that the foreseeability test, accompanied by the requirement for physical manifestation, sufficed.[153] In his forceful dissent, Justice Burke challenged this assumption. Liability based on foreseeability seemed to him too open-ended.[154] When the same court converted the *Dillon* guidelines to preconditions for recovery more than twenty years later, it did so explicitly in order to "avoid limitless liability out of all proportion to the degree of a defendant's negligence, and against which it is impossible to insure without imposing unacceptable costs on those among whom the risk is spread."[155]

The need to prevent abominable disproportion was regarded as the primary consideration to be weighed against "the impact of arbitrary lines [that] deny recovery to some victims whose [harms are] very real."[156] Similar reasoning was applied by other courts that adopted analogous prerequisites for liability.[157]

This line of reasoning has been supported in the academic literature. It is arguable, as one author noted, that "[a]llowing enormous damages to flow from merely negligent conduct runs the risk of unfairly penalizing a morally innocent defendant and of being perceived as grossly disproportionate to the defendant's fault."[158] One of the key concerns that ought to be addressed, when trying to

---

148. *Id.* at 27.

149. 512 U.S. 532, 551.

> A more significant problem is the prospect that allowing such suits can lead to unpredictable and nearly infinite liability for defendants . . . . This concern . . . is based upon the recognized possibility of *genuine* claims from the essentially infinite number of persons, in an infinite variety of situations, who might suffer real emotional harm as a result of a single instance of negligent conduct.

Id. at 552.

150. 45 U.S.C.A. §§ 51-60 (2004).

151. *Gottshall*, 512 U.S. at 557.

152. *Dillon*, 441 P.2d at 919.

153. *Id.* at 920-21.

154. *Id.* at 927-28 (Burke, J., dissenting).

155. Thing v. La Chusa, 771 P.2d 814, 826-27.

156. *Id.* at 827; *see also id.* at 821 (indicating that the post-*Dillon* decisions have not given adequate consideration to the "importance of avoiding the limitless exposure to liability that the pure foreseeability test of 'duty' would create.").

157. Brooks v. Decker, 516 A.2d 1380, 1382 (Pa. 1986); Bowen v. Lumbermens Mutual Casualty Co., 517 N.W.2d 432, 442-43.

158. Miller, *supra* note 127, at 19-20.

solve the problem of relational emotional harm, is the need to prevent a grossly disproportionate penalty for merely negligent conduct.[159] Even those who reject the exclusionary rule, with its impact or zone-of-danger exceptions, understand that some mechanism must be utilized to prevent abominable disproportion.[160]

The preceding analysis leads to the following conclusions. First, although courts and scholars differ on the proper method of limitation, no one has ever contended that any foreseeable emotional harm caused by the negligent infliction of injury upon a third party must be remedied.[161] Second, this consensus can be partly explained by the fear of abominable disproportion between the gravity of one's conduct and the severity of the ensuing sanction. In this article I do not attempt to determine where the line should be drawn. Perhaps there is no justifiable stopping-point beyond the zone of danger. Maybe a combination of seriousness and foreseeability is defensible. However, all possible resolutions are undeniably sensitive to the prospect of retributive injustice.

4. Addressing Several Criticisms

In the previous sections I argued that the courts have been reluctant to impose liability for relational losses at least partly due to the fear of an abominable disproportion between the gravity of the defendant's wrong and the severity of the sanction that it may entail. Critics may argue that the same fear exists with regard to *negligently* caused mass-disasters, such as an airplane crash, a collision at sea, or the collapse of a building. In cases of mass disaster the potential loss is indeed very large, and tort liability for the entire loss may be wholly out of proportion to the gravity of the wrong committed; nonetheless, the common law does not exclude liability for physical injuries caused in such incidents. Consequently, it cannot be said that prevention of abominable disproportion is regarded, in tort law, as a justification for exclusion of liability.[162]

This criticism of the argument that the law regarding relational loss was significantly influenced by the fear of abominable disproportion is flawed for several reasons. First, it is based on an assumption that is frequently incorrect. In many cases of mass disaster, the fear of extremely onerous liability does not

---

159. *Id.* at 3, 20, 35.

160. *Id.* at 35-36.

161. "Cause" is used here in its factual sense (i.e. cause-in-fact or *causa sine qua non*).

162. *Cf.* Gary T. Schwartz, *The Economic Loss Doctrine in American Tort Law: Assessing the Recent Experience*, *in* CIVIL LIABILITY FOR PURE ECONOMIC LOSS 103, 127 (Efstathios K. Banakas ed., 1996); Philip S. James, *The Fallacies of* Simpson v. Thomson, 34 MOD. L. REV. 149, 162 (1971); Stephen R. Perry, *Protected Interests and Undertakings in the Law of Negligence*, 42 U. TORONTO L.J. 247, 263 (1992); Robert L. Rabin, *Tort Recovery for Negligently Inflicted Economic Loss: A Reassessment*, 37 STAN. L. REV. 1513, 1529 n. 52, 1532 (1985); Kelly M. Hnatt, Note, *Purely Economic Loss: A Standard for Recovery*, 73 IOWA L. REV. 1181, 1210 (1988).

arise because tort recovery is limited by a specific statute, and the potential injurer is fully insured. This is so with regard to aerial accidents,[163] marine accidents,[164] oil pollutions,[165] and even nuclear incidents.[166] In cases of mass disaster where no statutory limits exist, as in conflagration, the common law of torts has been adapted to prevent abominable disproportion between the extent of liability and the gravity of the wrong.[167]

Second, even if recovery for physical injuries were not legally limited, the physical impact of an accident—as opposed to its economic ramifications—would usually be determinate.[168] The ability to assess in advance the magnitude of the loss makes liability insurance far more feasible, and this militates against abominable disproportion between the burden actually incurred by the injurer and the gravity of his or her conduct. Relational losses, on the other hand, may ripple indeterminately, making full liability insurance either infeasible or extremely expensive. In either case, an injurer may have to bear a grossly disproportionate sanction.

Third, proportionality is not the primary concern of tort law. On the contrary, it is usually overshadowed by other legitimate goals of civil liability. It is arguable that the non-retributive considerations that support liability in cases of mass-disaster are much stronger than the considerations that support liability for relational losses, both economic or emotional. For example, one may say that life and bodily-integrity are a lot more valuable than purely economic or emotional interests from a social perspective, and, therefore, merit more comprehensive protection by the law.[169] It may, thus, be justifiable to render them more extensive protection, even if the outcome is extremely unjust from a retributive perspective.

Critics of my fundamental argument may also argue that while allowing recovery for relational losses may give rise to disproportion between the extent of liability and the gravity of the defendant's conduct, exclusion of liability will definitely result in disproportion between the burden incurred by the relational victim and her complete innocence.[170] However, from a retributive perspective the fairness of imposing a certain sanction on a certain person must be determined by a single factor, namely the gravity of the wrong for which the

---

163. *Supra* note 108.

164. *Supra* notes 106-07 and accompanying text.

165. 33 U.S.C. § 2704 (2005).

166. 42 U.S.C. § 2210 (2005).

167. *E.g.*, Ryan v. New York Central R.R. Co., 35 N.Y. 210, 216-17 (1866) (finding that a person who negligently starts a fire is not liable to his neighbors for all property damage resulting from the fire).

168. For example, an airline, or an airplane manufacturer, can predict with accuracy the number of casualties in case of an airplane crash; similarly, a building contractor can foresee the number of victims in case of a collapse.

169. Patrick S. Atiyah, Negligence and Economic Loss, 83 L.Q. REV. 248, 267 (1967); Fleming James, Jr., *Limitations on Liability for Economic Loss Caused by Negligence: A Pragmatic Appraisal*, 25 VAND. L. REV. 43, 54 n.45 (1972).

170. Kinsman Transit Co. v. City of Buffalo, 388 F.2d 821, 823.

sanction is imposed.[171] If the severity of a certain sanction does not fit the gravity of the wrong, retributive justice mandates not to impose it, regardless of any other retributive injustice that may ensue:[172] one cannot resolve one retributive injustice by actively creating another. The status of the relational victim will be similar to that of a person injured by non-negligent conduct of a doctor, a hurricane or lightning.[173] If it is believed that her suffering merits reparation, on retributive or other grounds, other mechanisms should be employed to deal with it (such as social security or mandatory insurance).

Furthermore, it seems that courts do not consider the grievance that may result from exclusion of liability for relational losses to be as appalling as the one that may arise from imposition of liability. If the exclusionary rule were abolished, a person who deviated from the objective standard of care only slightly and absentmindedly could nonetheless face a staggering, and abominably disproportionate, liability. On the other hand, exclusion of liability makes the innocent victim incur only a small fraction of the aggregate loss.[174]

Lastly, no argument based on the injurer's fault and the victim's innocence can ignore two common traits of tort settings. First, the defendant is very often the "innocent" employer of the actual wrongdoer. Second, at least in certain cases of relational economic loss, the victim may not be deemed wholly innocent, because she could easily have protected herself from the harm through first-party insurance, contractual provisions, *ex ante* precautions, or *ex post* measures of mitigation.[175]

B. Punishing the Victim for Illicit Conduct

1. The Notion of Ex Turpi Causa Non Oritur Actio

a. Introduction

In the seminal case of *Holman v. Johnson,*[176] Lord Mansfield

---

171. *See* sources cited *supra* note 20.

172. One may think of another example of this principle. When a murderer is incarcerated, his family may lose a breadwinner and incur serious financial, and emotional, harm. If the family members are innocent, the sanction imposed on them will probably be considered too harsh. From a retributive perspective, this does not make the criminal less deserving of his punishment.

173. In other words, the relational victim is not entitled to recover for her loss, despite the fact that she may have been completely innocent.

174. *Cf.* Amaya v. Home Ice, Fuel & Supply Co., 379 P.2d 513, 525

> It begs the question to argue that "If the loss is out of all proportion to the defendant's fault, it can be no less out of proportion to the plaintiff's innocence." That obvious truism could be urged by every person who might adversely feel some lingering effect of the defendant's conduct, and we would then be thrown back into the fantastic realm of infinite liability.

Id. (citations omitted).

175. *See* Perry, *supra* note 78, at 757-58, 766-68, 774.

176. (1775) 98 Eng. Rep. 1120, 1121 (K.B).

acknowledged, for the first time in English jurisprudence, the principle later known as *ex turpi causa non oritur actio*, or the illegality defense.[177] In his words, "[n]o court will lend its aid to a man who founds his cause of action upon an immoral or an illegal act."[178] Lord Mansfield made this statement in a contractual context, and its applicability within the law of contracts is now uncontested. It is universally accepted that a contract whose formation or performance is criminal, immoral, or otherwise counter to public policy should be deemed invalid.[179]

In many jurisdictions the doctrine of *ex turpi causa* eventually extended to tort law. Clearly, whenever an action in tort is based on a breach of an illegal contract, such action must fail.[180] But the abstract idea that enforceable legal rights should not arise from unlawful conduct does not, and should not, depend upon the exact manifestation of unlawfulness.[181] As noted by Lord Diplock forty years ago:

> [E]x turpi causa non oritur actio, is concerned not specifically with the lawfulness of contracts but generally with the enforcement of rights by the courts, whether or not such rights arise under contract. All that the rule means is that the courts will not enforce a right . . . if the right arises out of an act committed by the person asserting the right . . . which is regarded by the court as sufficiently anti-social to justify the court's refusing to enforce that right.[182]

The line between the contributory negligence defense and the *ex turpi causa* doctrine is not always drawn correctly. The former focuses on the fact that the plaintiff's conduct endangered herself,[183] whereas the latter focuses on the unlawfulness of the plaintiff's conduct in relation to others. The

---

177. *Id.*

178. *Id.* Cited with approval in Higgins v. McCrea, 116 U.S. 671, 685-86 (1886).

179. W.V.H. ROGERS, WINFIELD AND JOLOWICZ ON TORT 863 (16th ed. 2002); RESTATEMENT (SECOND) OF CONTRACTS § 178 (1986).

180. MARGARET BRAZIER & JOHN MURPHY, THE LAW OF TORTS 106 (10th ed. 1999).

181. *See, e.g.*, Hall v. Corcoran, 107 Mass. 251, 253 (1871), in which Justice Gray states that

> [C]ourts of justice will not assist a person who has participated in a transaction forbidden by statute to assert rights growing out of it, or to relieve himself from the consequences of his own illegal act. Whether the form of the action is in contract or in tort, the test in each case is, whether, when all the facts are disclosed, the action appears to be founded in a violation of law, in which the plaintiff has taken part.

See also Miller v. Bennett, 56 S.E.2d 217, 219 (Va. 1949) ("The principle applies to civil actions, whether based on tort or contract.").

182. Hardy v. Motor Insurers' Bureau, [1964] 2 All E.R. 742, 750-51 (Eng. C.A.); *see also* Hall v. Woolston Hall Leisure Ltd., [2001] 1 W.L.R. 225, 247 (Eng. C.A.); Standard Chartered Bank v. Pakistan Nat'l Shipping Corp., [2000] 1 Lloyd's Rep. 218, 232 (Eng. C.A.).

183. RESTATEMENT (SECOND) OF TORTS § 463 (1986). "Contributory negligence is conduct on the part of the plaintiff which falls below the standard to which he should conform for his own protection . . . ." *Id.*; *see also id.* § 466(b) (defining contributory negligence as "conduct which . . . falls short of the standard to which the reasonable man should conform in order to protect himself from harm").

justification for the contributory negligence doctrine is that one cannot sue another for the materialization of risk created by oneself.[184] As such, contributory negligence and its modern variation—comparative negligence—clearly comply with the notion of corrective justice.[185] The defendant must compensate the plaintiff for any harm caused by his wrongful conduct, while the plaintiff must bear her loss or a certain fraction of it if it was caused by her own wrongful conduct.

Although the *ex turpi causa* defense is normally contingent on the causal connection between the plaintiff's conduct and her harm,[186] it is much less interested in the fact that the plaintiff's conduct contributed to her harm and in the extent of such contribution than in the fact that the plaintiff's conduct was offensive to the public at large.[187] Accordingly, *ex turpi causa* cannot be

---

184. Smith v. Smith, 19 Mass. 621, 623 (1824); Whirley v. Whiteman, 38 Tenn. 610, 619 (1858) ("[I]f by ordinary care and prudence plaintiff might have avoided her harm, she must be regarded as the author of her own misfortune."); Francis H. Bohlen, *Contributory Negligence*, 21 HARV. L. REV. 233, 256 (1908).

> [W]here. . . the defendant's delinquency would have caused no harm to the plaintiff save for his own misconduct in not caring for himself, there is no reason that the law should regard one as the delinquent rather than the other. There is no reason to throw upon the one rather than the other the burden of preventing an accident actually preventable by proper care on the part of either, or of answering for the ensuing harm. It is for this reason, and because the law will not aid a plaintiff who having the power and consequent duty to protect himself has failed to do so . . . that the defendant is relieved from liability by the plaintiff's contributory negligence.

Id.

185. For further discussion of the substitution in negligence actions of the principles of damage apportionment embodied by the comparative negligence doctrine for the "all or nothing" approach of the contributory negligence doctrine see Thomas R. Trenkner, Annotation, *Modern Development of Comparative Negligence Doctrine Having Applicability to Negligence Actions Generally*, 78 A.L.R.3d 339 (1977). *See also* RESTATEMENT (SECOND) OF TORTS § 467, Special Note (1965) (explaining the notion of comparative negligence without using the actual term). For a Tennessee perspective see Brian P. Dunigan & Jerry J. Phillips, *Comparative Fault in Tennessee: Where Are We Going, and Why Are We in this Handbasket?*, 67 TENN. L. REV. 765 (2000).

186. *See, e.g.*, Havis v. Iacovetto, 250 P.2d 128, 130 (Colo. 1952) (finding that plaintiff's illegality does not bar recovery because it was not a proximate cause of the injury); Johnson v. Thompson, 143 S.E.2d 51, 54 (Ga. 1965) (finding that plaintiff's gambling does not bar recovery when there is no causal connection); Martinez v. Rein, 146 So. 787, 787 (La. Ct. App. 1933) (finding that violation of a city ordinance that did not lead to the injury was not a bar to recovery); Janusis v. Long, 188 N.E. 228, 232 (Mass. 1933) (finding that status as an illegal alien is not a bar to recovery); Manning v. Noa, 76 N.W.2d 75, 76-77 (Mich. 1956) (finding that playing bingo was not the proximate cause of plaintiff's injury, and was, therefore not a bar to recovery); Meador v. Hotel Grover, 9 So. 2d 782, 785 (Miss. 1942) (finding that the violation of the hotel statute did not directly lead to the injury, so no bar to recovery); Curry v. Vesely, 348 P.2d 490, 491 (N.M. 1960) (noting that there must be a causal connection between the unlawful assembly and the injury suffered).

187. Barker v. Kallash, 468 N.E.2d 39, 40 (N.Y. 1984) (explaining that contributory negligence bars recovery because the plaintiff contributed to his injury, whereas the notion of *ex*

regarded as a manifestation of corrective justice. The Supreme Court of Canada correctly observed that "[l]iability for tort arises out of the relationship between the alleged tortfeasor and the injured claimant,"[188] whereas, the power of the court to deny recovery under *ex turpi causa* "represents concerns independent of this relationship."[189] Based on the foregoing, the argument that the *ex turpi causa* defense was implicitly abolished by the legislature when it replaced the traditional rule of contributory negligence with the more flexible principle of comparative negligence is unfounded.[190] The above argument is an inevitable outcome of a misconception of *ex turpi causa*.[191]

Traditionally, *ex turpi causa* functioned as a bar to recovery for any harm incurred by the plaintiff during, or following, the performance of any criminal offense, regardless of its severity.[192] The underlying rationale for this practice was that those who transgress the criminal code should not receive aid from the judiciary.[193] In other words, a person who breaks the law does not deserve to be protected by the law against the consequences of her own conduct.[194] A few

---

turpi causa bars recovery because public policy denies judicial relief to those injured while committing a serious crime); *see also id.* at 43.

188. Hall v. Herbert, [1993] 101 D.L.R. (4th) 129, 169.

189. *Id.*; *see also id.* at 170 (noting that "the legality or morality of the plaintiff's conduct is an extrinsic consideration" in a bipolar corrective pattern).

190. This argument was brought up, for example, in *Hall*, [1993] 101 D.L.R. (4th) at 154-55 (Cory J., concurring); Lewis v. Sayers, [1970] 13 D.L.R. (3d) 543, 550.

191. *Barker*, 468 N.E.2d at 43-44 (stating that the legislative abolition of contributory negligence and adoption of comparative negligence "has no application to the rule precluding a plaintiff from recovering for injuries sustained as a direct result of his own illegal conduct of a serious nature involving risk of physical harm.").

192. *See, e.g.*, Newton v. Illinois Oil Co., 147 N.E. 465, 468 (Ill. 1925) (finding that plaintiff was not entitled to recover for the wrongful death of his son at work because he permitted his son to be employed contrary to the prevailing child labor laws); Fristoe v. Boedeker, 194 Ill. App. 52, 57 (1915) (finding that plaintiff was not entitled to compensation for severe injuries incurred while he was hunting on private land without the owner's permission, when such sanctions were a misdemeanor under State law); Lencioni v. Long, 361 P.2d 455, 457-58 (Mont. 1961) (finding that plaintiff was not allowed to recover for injuries sustained during his work at a brothel, the operation of which was illegal under State law); Curtis v. Murphy, 63 Wis. 4, 7-8 (1885) (finding that plaintiff was not allowed to recover from a hotel whose clerk stole plaintiff's money because he checked into that hotel to have illicit intercourse with a prostitute).

193. Bonnier v. Chicago B&Q R.R. Co., 113 N.E.2d 615, 622 (Ill. App. Ct. 1953).

194. A second frequently used justification for the traditional version of *ex turpi causa* was that "one may not profit from one's own wrongdoing." Manning v. Brown, 689 N.E.2d 1382, 1384 (N.Y. 1997); *see* Wiley v. County of San Diego, 966 P.2d 983, 990 (Cal. 1998). However, this rationale seems dubious given that in most cases exclusion of liability does not, in fact, deprive the plaintiff from any profits of her wrongdoing. *See* Gail D. Hollister, *Tort Suits for Injuries Sustained during Illegal Abortions: The Effects of Judicial Bias*, 45 VILL. L. REV. 387, 392 (2000); Joseph H. King, *Outlaws and Outlier Doctrines: The Serious Misconduct Bar in Tort Law*, 43 WM. & MARY L. REV. 1011, 1044 (2002); *see also* THE LAW COMMISSION, CONSULTATION PAPER NO. 160: THE ILLEGALITY DEFENCE IN TORTS 77 (2002) [hereinafter THE ILLEGALITY DEFENCE IN TORTS]. "[I]n most tort cases the rationale of preventing a claimant

notable examples demonstrate the harshness of this perception. In *Bosworth v. Swansey,*[195] the plaintiff was injured because of a defect in the road for which defendant was responsible.[196] However, because the accident occurred on Sunday, and traveling on Sunday for secular purposes was prohibited by criminal statute, the court denied recovery.[197] In *Heland v. Lowell,*[198] the plaintiff was injured while riding his horse "at a rate faster than a walk" across a bridge maintained by the defendant in violation of a penal by-law.[199] Following the rationale in *Bosworth*, the court stated that "when a plaintiff's own unlawful act concurs in causing the damage that he complains of, he cannot recover compensation for such damage."[200]

Denial of recovery under the traditional understanding of *ex turpi causa* was a sanction attached to *every* crime, beyond any sanction prescribed by criminal law. It served as an integral part of criminal punishment, although it was formally imposed by tort law. So its justifications had to be similar to those of criminal punishment, i.e. retribution and deterrence.[201] In time, courts came to understand that the draconian application of *ex turpi causa* was inconsistent with at least one of its apparent justifications: denial of recovery for any crime-related harm, regardless of the gravity of the crime and the severity of the actual or expected punishment under criminal law, seems unfair

---

profiting from his or her own wrongdoing will not justify the application of the illegality doctrine." *Id.*

A third justification for the traditional rule is that it deters potential criminals. For several persuasive criticisms of this reasoning see King, at 1045-46 (observing, among other things, that most criminals do not expect to be negligently injured, and, therefore, would seldom be influenced by the illegality defense; that whenever there is an apparent risk of personal injury it generates, by itself, a significant deterrent; and that in any case *ex turpi causa* can hardly augment the deterrent effect of criminal sanctions).

195. 51 Mass. 363 (1845).

196. *Id.* at 365.

197. *Id.* at 365-66; *accord* Hinckley v. Penobscot, 42 Me. 89, 93 (1856); Read v. Boston & Albany R.R. Co., 4 N.E. 227 (Mass. 1885) (finding that plaintiff locomotive engineer was barred from recovering for injuries caused by defective track because he was running an unauthorized Sunday train); Lyons v. Desotelle, 124 Mass. 387, 390 (1878) (finding that unlawful traveling on Sunday bars recovery for harm resulting from defendant's negligence); Smith v. Boston & Maine R.R., 120 Mass. 490, 493 (1876); Gregg v. Wyman, 58 Mass. 322, 325-26 (1849) (finding that plaintiff could not recover for injury to horse rented in violation of a statute prohibiting horse rental on Sunday). For further discussion of cases denying recovery for injuries sustained while violating Sunday laws see THOMAS M. COOLEY, A TREATISE ON THE LAW OF TORTS OR THE WRONGS WHICH ARISE INDEPENDENT OF CONTRACTS 152-55 (1880).

198. 85 Mass. 407 (1862).

199. *Id.* at 408.

200. *Id.*; *see* Tuttle v. Lawrence, 119 Mass. 276, 278 (1876); *cf.* Dudley v. Northampton St. Ry. Co., 89 N.E. 25, 28 (Mass. 1909) (finding that a plaintiff injured while driving without Massachusetts license could not recover).

201. *Cf.* Harold S. Davis, *The Plaintiff's Illegal Act as a Defense in Actions of Tort*, 18 HARV. L. REV. 505, 513 (1905) (explaining that the doctrine of *ex turpi causa* operates "in the nature of a punishment for the plaintiff's wrongdoing.").

from a retributive perspective.[202]

At present, most jurisdictions do not deny recovery for any harm or injury suffered by an offender during or in consequence of her own criminal conduct.[203] The old doctrine applies in limited circumstances only. The extent of its application varies among different jurisdictions, but it does not usually transcend the following archetypal categories: (1) cases in which plaintiff's conduct is regarded as extremely grievous; (2) cases in which denial of recovery is required to preserve criminal justice; (3) cases in which plaintiff and defendant participated in the same illegal conduct. These categories, which may somewhat overlap, are discussed in detail below. It will be shown that denial of liability in at least two of them may have a retributive foundation.

b. Punishing Extremely Grievous Conduct

In recent decades, courts employed *ex turpi causa* to exclude recovery by plaintiffs whose conduct was perceived to be extremely grievous.[204] For example, in *Barker v. Kallash*[205] a teenager was injured when a pipe-bomb that he was constructing using material supplied by the defendants exploded. The

---

202. Hall v. Herbert, [1993] 101 D.L.R. (4th) 129, 152 (Cory J., concurring). "There does not seem to be any rational basis for a court to impose an additional sanction upon a convicted person by denying what may well be fair and just compensation for injuries received as a result of a tortious act." *Id. See also* THE LAW COMMISSION, CONSULTATION PAPER NO. 154: ILLEGAL TRANSACTIONS: THE EFFECT OF ILLEGALITY ON CONTRACTS AND TRUSTS 106 (1999) [hereinafter THE EFFECT OF ILLEGALITY ON CONTRACTS].

> The simple refusal of civil relief is generally a very arbitrary and blunt method of meting out punishment, since the penalty is not in any way tailored to fit the illegality involved. And clearly there will be a risk of "double punishment" where the plaintiff has already been convicted of a criminal offence or made to pay damages for a legal wrong in respect of the same conduct. . . . [I]f a sanction has already been imposed on the plaintiff in respect of his or her unlawful conduct, then the additional denial of civil relief might be regarded as unduly harsh.

Id.; Ernest J. Weinrib, *Illegality as a Tort Defence*, 26 U. TORONTO L.J. 28, 45-46 (1976) (setting forth that if criminal penalty is retributively just, an additional sanction would make the aggregate penalty undue).

203. *See, e.g.*, Barker v. Kallash, 468 N.E.2d 39, 44 (Jasen J., concurring) ("This so-called 'outlaw' doctrine of tort law -- i.e., depriving a violator of the law of any rights against a tort-feasor -- has long since been discarded by most, if not all, American jurisdictions."); *cf.* THE ILLEGALITY DEFENCE IN TORTS, *supra* note 194, at 9 ("[T]he law does not allow even a criminal who has committed a serious offence to be deprived of all his or her rights under either the civil or criminal law. This would amount to outlawry, and this has quite clearly, and in our view rightly, been rejected by the courts.") (discussing UK law).

204. The examples that follow in the text are taken from American case-law. However, similar rhetoric may be found in non-American jurisdictions. *See, e.g., Standard Chartered Bank v. Pakistan Nat'l Shipping Corp.*, [1998] 1 Lloyd's Rep. 684, 705-06 (Eng.). "Whatever theory founds a defence of ex turpi, the defendant must establish [among other things] that the plaintiff's conduct is so clearly reprehensible as to justify its condemnation by the Court." *Id.* at 705.

205. 468 N.E.2d 39.

Court of Appeals of New York reasoned that, when the plaintiff's injury is a direct result of his knowing and intentional participation in a criminal act, he cannot seek compensation for the loss when the criminal act is judged to be *so serious an offense* as to warrant denial of recovery.[206] Justice Jasen opined that the court would bar recovery only if the plaintiff's violation of the law was either gravely immoral or grievously injurious to the public interests, as is the case with rape or arson.[207] In this particular case, "the plaintiff's grievous criminal conduct . . . was so plainly violative of paramount public safety interests, [that] the public policy of this State dictates that recovery be denied."[208] In subsequent cases, the Court of Appeals adhered to the rule that liability will be denied only if the plaintiff's conduct constitutes such a serious offense that public policy precludes her recovery.[209] This condition was usually met when the plaintiff knowingly and significantly jeopardized the lives of other people, as in the *Barker* case.[210]

Similarly, the Supreme Court of Alaska concluded in *Ardinger v. Hummell*[211] that the principle of no-recovery applied only in cases involving serious criminal conduct that intentionally threatened the safety of others, such as homicide, rape, and arson.[212] The court reasoned that the harsh sanction of no-recovery should be reserved for extreme circumstances.[213] Consequently, the court rejected the argument that a person driving a car without its owner's permission, in violation of an Alaska statute, should not be allowed to recover for injuries incurred in the course of such driving.[214]

In *Lewis v. Miller*,[215] the Superior Court of Pennsylvania found that a person whose violation of the law shows conscious indifference to her own safety and to public safety generally cannot recover for her consequent injuries.[216] Denial of recovery requires proof of "wantonness," at a minimum.[217] In *Lewis*, the claim was dismissed because the decedent was

---

206. *Id.* at 41 (emphasis added). The court reiterates the same idea throughout the opinion. *Id.* at 41-44; *see also id.* at 45 (Jasen J., concurring) (arguing that recovery should be denied only where plaintiff's conduct was "so egregious an offense that permitting recovery would be inimical to the public interest").

207. *Id.* at 45 (Jasen J., concurring).

208. *Id.* For further examples see King, *supra* note 194, at 1025-27.

209. Manning v. Brown, 689 N.E.2d 1382, 1384 (N.Y. 1997). *See also* Guadamud v. Dentsply Int'l, Inc., 20 F. Supp. 2d 433, 436 (1998) (applying New York law).

210. *Manning*, 689 N.E.2d at 1384-85 (finding that reckless unauthorized joyriding without a driver's license constitutes a serious violation).

211. 982 P.2d 727 (Alaska 1999).

212. *Id.* at 736.

213. *Id.*

214. *Id.* "The offense on which [defendant] relies cannot reasonably be equated with crimes such as homicide, rape, and arson for purposes of barring recovery on public policy grounds.... Such a violation does not represent the level of serious criminal conduct generally necessary to bar recovery." *Id.*

215. 543 A.2d 590 (Pa. Super. Ct. 1988).

216. *Id.* at 592-93.

217. *Id.* at 592.

found to have engaged in wanton conduct, namely participating in a "drag race" on a dangerous blind curve while intoxicated.[218]

Finally, in *Oden v. Pepsi Cola Bottling Co.*,[219] the Supreme Court of Alabama found that the law precludes "any action seeking damages based on injuries that were a direct result of the injured party's knowing and intentional participation in a crime *involving moral turpitude*."[220] The same court explained, in a subsequent case, that a crime involving moral turpitude exhibits "an inherent quality of baseness, vileness, or depravity in regard to the duties one owes to society."[221] *Ex turpi causa* was reserved, at least in theory, for extreme cases.

The court in *Oden* denied recovery to the parents of an adolescent who was killed when a soft-drink vending machine fell on him while he attempted to steal drink cans.[222] This case does not seem as heinous as *Barker* or as aggravating as *Lewis*. Those who believe that *ex turpi causa* must be reserved for especially heinous cases may, consequently, consider its application in *Oden* inappropriate; however, the significance of *Oden* lies not in the specific application of the archaic doctrine, but in the determination of its limited role in tort law. The court applied this limitation more convincingly two years later, when it found that knowingly being a passenger in a car driven by another, without license or permission, does not involve base, vile, or depraved conduct. Hence, the passenger may recover for injuries resulting from the negligence of third parties.[223]

The view that denial of recovery on grounds of illegality ought to be reserved for the most grievous offenses appears to be motivated by the notion of retributive justice.[224] This hypothesis is supported by two possible lines of argument. First, it may be said that the *judicial restriction of the traditional doctrine* was aimed at preventing abominable disproportion between the severity of the sanction imposed on the offender and the gravity of her conduct. To understand this contention one must recall that traditionally *ex turpi causa* was used to deny recovery for any crime-related harm. This is the positive baseline.

Assume, for the moment, that the criminal offender is punished by criminal law. If the criminal penalty is determined considering the need for just desert, an additional sanction in the form of no-recovery for personal injuries in tort may produce disproportion between the crime and the punishment. Indeed, this type of injustice may occur regardless of the gravity of the offense. This perception might call for a total abolition of *ex turpi causa*.

However, as courts seem reluctant to abolish well established and time-

---

218. *Id.*

219. 621 So. 2d 953 (Ala. 1993).

220. *Id.* at 955 (emphasis added).

221. Lemond Constr. Co. v. Wheeler, 669 So. 2d 855, 861 (Ala. 1995).

222. *Oden*, 621 So. 2d at 954.

223. *Lemond Constr. Co.*, 669 So. 2d at 861.

224. Deterrence does not seem to be an equally persuasive explanation. *See supra* note 194.

honored common law doctrines, it is more likely that highly problematic doctrines will be restricted rather than eliminated by the judiciary. It is, therefore, conceivable that while hesitant to abolish the *ex turpi causa* defense, courts will limit its application to extreme cases in order to ensure that the discrepancy between the gravity of the wrong and the severity of combined sanction is not too significant. The more grievous the conduct, the less likely it is that depriving the offender of her civil remedy will result in considerable (not to say abominable) disproportion. The more heinous the crime, the more likely it is that denial of tort recovery will not violate the principles of proportionality at all.[225]

The alternative line of argument is that, regardless of the historical baseline, denial of tort liability when the plaintiff's illegal conduct is "serious" may be justified within the prevention of abominable disproportion paradigm. This is so only if a second factor is introduced into the equation: the lenience of the actual or expected criminal sanction. The *ex turpi causa* defense can be justified as preventing abominable disproportion if without its application the plaintiff may escape any penalty, or face a relatively lenient sanction, for a serious wrongdoing. This may happen when the plaintiff's crime is left with impunity, or when her conduct is outrageous but not prohibited by penal law. Private law, tort law in particular, will not always be capable of fixing such abominable disproportion. It can do so only when the unpunished offender seeks the aid of the court in a civil action linked with her illicit conduct. On the other hand, given its commitment to corrective justice, tort law does not attempt to fix regular discrepancies. Courts concede corrective justice for the sake of proportionality only in rare cases.

This line of argument may explain cases like *Barker*, *Manning*, *Lewis*, *Oden*, and many others, in which the plaintiffs committed serious crimes but were not indicted in criminal courts. The plaintiff in *Barker* could not be held criminally responsible for his conduct because he was a minor at the time of the incident.[226] In *Manning*, the plaintiff plead guilty to charges relating to her crime, but she later withdrew her plea and was never prosecuted, perhaps due to the injuries she sustained.[227] The victims in *Lewis* and *Oden* were killed while participating in "serious" criminal activities, so they could not be indicted at all.[228] It is true that the courts in all four cases emphasized the seriousness of the conduct and neglected the absence of a criminal sanction, but perhaps they might have changed their decision if the offenders had been convicted and incarcerated for their grievous crimes.

---

225. One should bear in mind that the principles of ordinal and cardinal proportionality do not dictate a precise sanction for a given wrong. They allow the imposition of various sanctions within a certain range.

226. Barker v. Kallash, 468 N.E.2d 39, 43 (N.Y. 1984).

227. Manning v. Brown, 689 N.E.2d 1382, 1383 (N.Y. 1997).

228. Then again, it may be argued that in these two cases a capital punishment was too harsh.

c. Preserving Criminal Justice

In many jurisdictions the primary function of *ex turpi causa* is to preserve criminal justice as prescribed by criminal law.[229] It prevents tort liability from overriding the criminal sanction, either directly, by compensating the offender for the criminal penalty she had to endure, or indirectly, by allowing the offender to retain the benefits of her crime. The Supreme Court of Canada explicitly limited the application of the *ex turpi causa* defense in this manner. In *Hall v. Herbert*,[230] Justice McLachlin, with whom Justices La Forest, L'Heureux-Dubé and Iaccobucci concurred, opined that:

> [C]ourts should be allowed to bar recovery in tort on the ground of the plaintiff's immoral or illegal conduct only in very limited circumstances. The basis of this power, as I see it, lies in the duty of the courts to preserve the integrity of the legal system, and is exercisable only where this concern is in issue. This concern is in issue where a damage award in a civil suit would, in effect, allow a person to profit from illegal or wrongful conduct, or would permit to an evasion or rebate of a penalty prescribed by the criminal law. The idea common to these instances is that the law refuses to give by its right hand what it takes away by its left hand.[231]

Although there are no similar clear-cut definitions of the role of *ex turpi causa* in other jurisdictions, this doctrine is frequently used to vindicate criminal justice in contemporary Anglo-American case law, as will be demonstrated in the following paragraphs.

Clearly, one of the most prominent functions of *ex turpi causa*, in modern times, is to prevent compensation for a criminal penalty that was rightly imposed, and for criminal proceedings that were justifiably initiated.[232] For example, in *Braunstein v. Jason Tarantella, Inc.*,[233] the producers of an obscene film sued its distributors, *inter alia*, for negligence in distributing the film in a way that subjected the producers to criminal prosecution and for fraud in misrepresenting the number of places where such a film could be legally shown.[234] The Appellate Division of the Supreme Court of New York found that since the producers themselves committed the crime of obscenity in New York, they could not sue the distributors for making it possible for the producers to be caught and prosecuted.[235] To allow recovery would, in essence,

---

229. For recent reiterations of this idea in Canada *see, e.g.*, National Bank Financial Ltd. v. Potter, No. 174293, 36 A.C.W.S. (3d) (N.S.S.Ct. Jan. 12, 2005) *available at* 2005 A.C.W.S.J. LEXIS 75, *22-23; X. v. M. (R.D.), No. 5012573, 134 A.C.W.S. (3d) (B.C.S.Ct. Oct. 1, 2004) *available at* 2004 A.C.W.S.J. 13584, **178-91.

230. [1993] 101 D.L.R. (4th) 129.

231. *Id.* at 160.

232. *See* Weinrib, *supra* note 202, at 50-54.

233. 450 N.Y.S.2d 862 (1982).

234. *Id.* at 864.

235. *Id.* at 866.

permit the producers to recover the fine they paid upon their conviction.[236] Civil liability would offset a duly imposed criminal sanction.

Similarly, it was held that a man convicted of manslaughter for the fatal shooting of another could not recover damages from the manufacturer and the seller of the shotgun for the legal consequences of his deeds;[237] a woman who was convicted for murdering her husband could not recover damages from her psychiatrist on the theory that he negligently failed to prevent her from committing the murder;[238] and, a man, convicted and incarcerated for kidnapping, raping, and assaulting a woman while intoxicated, could not recover damages from a bar and a bartender on the theory that they caused his imprisonment by negligently serving him alcoholic beverages after he became intoxicated.[239]

An English counterpart is *Clunis v. Camden & Islington Health Authority*.[240] In *Clunis*, the plaintiff was convicted of manslaughter, and he argued that the defendant's failure to provide him with proper mental health care led to the commission of the crime and his subsequent conviction.[241] The plaintiff's claim was dismissed. Had he been allowed to recover, his punishment might have been diminished because of the monetary amends.[242] The court's denial of liability thwarted the plaintiff's clear attempt to shirk responsibility and preserved criminal justice.[243]

Denial of compensation for criminal penalty may be justified, at least in part, by the notion of just desert. If tort law allowed the offender to receive compensation for being tried and punished it would counteract criminal justice;

---

236. *Id.* Prior to dismissing the tort claims the Appellate Division rejected the producers' contractual claim for the proceeds of distributing the film in New York and in other states, saying that to allow recovery "would be to call upon this court to serve as pay master of the wages of crime." *Id.* at 865.

237. Adkinson v. Rossi Arms Co., 659 P.2d 1236, 1239-40 (Alaska 1983). The court's denial of liability was phrased as a no-duty decision, but the formal pigeon hole into which the case was classified is of little relevance here. The court stated that "allowing a criminal . . . who has been convicted of an intentional killing, to impose liability on others for the consequences of his own anti-social conduct runs counter to basic values underlying our criminal justice system." *Id.* at 1240.

238. Cole v. Taylor, 301 N.W.2d 766, 768 (Iowa 1981); *see also* Glazier v. Lee, 429 N.W.2d 857, 860 (Mich. Ct. App. 1988) (finding that a man who killed his girlfriend could not recover from his psychologist for negligently failing to prevent him from committing voluntary manslaughter).

239. Lord v. Fogcutter Bar, 813 P.2d 660, 662-63 (Alaska 1991).

240. [1998] Q.B. 978.

241. *Id.* at 984-85.

242. *Id.* at 990, 993.

243. *See also* Colburn v. Patmore, [1834] 149 Eng. Rep. 999, 1003 (Exch. Div.). In that case, plaintiff publisher, who had been convicted and fined for publishing a criminal libel, sued the editor who had published the material without plaintiff's knowledge. *Id.* at 999-1000. The plaintiff sought to force the editor to compensate him for the amount of the criminal fine the plaintiff had been forced to pay because of the editor's acts and recovery was denied. *Id.* at 999. Allowing recovery in such a case would have stultified the effect of the criminal law.

the law would give with one hand (tort law) what had been taken with the other (criminal law). As observed by Justice Denning in *Askey v. Golden Wine Co.*,[244] the objects of criminal law "would be nullified if the offender could recover the amount of the fine and costs from another by process of the civil courts."[245] Denial of recovery maintains criminal justice, thereby advancing its primary goals of retribution and deterrence.

A second function of *ex turpi causa* is to prevent the criminal offender from acquiring the benefits of her illegal conduct, thereby sustaining the severity of the criminal punishment. In fact, this function is the only one that is truly consistent with one of the most frequently cited justifications for the *ex turpi causa* doctrine in Anglo-American jurisprudence: "[O]ne may not profit from one's own wrongdoing."[246] For example, in *Mettes v. Quinn*,[247] the plaintiff sued her former attorney, claiming that his negligent advice caused her fraudulent acts to be uncovered, depriving her of the expected benefit of the fraud.[248] It was held that a court "will not aid a fraudfeasor who invokes the court's jurisdiction to profit from his own fraud by recovering damages."[249]

An English parallel is *Thackwell v. Barclays Bank Ltd.*,[250] in which the plaintiff claimed reimbursement of sums that represented the proceeds of a fraud in which he participated.[251] Recovery was denied on grounds of illegality. The court observed that if it allowed recovery, it would "be indirectly assisting the commission of a crime."[252] Note that the *ex turpi causa* defense is used not only to bar an entire claim, but also to disallow compensation under particular heads of damages, if only those reflect the lost proceeds of an illegal business, occupation, or activity.[253]

---

244. [1948] 2 All E.R. 35 (K.B.).

245. *Id.* at 38.

246. *See* King, *supra* note 194 (quoting Manning v. Brown, 689 N.E.2d 1382, 1384 (N.Y. 1997).

247. 411 N.E.2d 549 (Ill. App. Ct. 1980).

248. *Id.* at 550-51.

249. *Id.* at 551; *accord* Lewis v. Brannen, 65 S.E. 189, 191 (Ga. Ct. App. 1909) (denying claim by plaintiff who lost the proceeds of selling a medicine due to the negligent supply of one of its ingredients by the defendant because he sold the drug without the necessary license); J.R. Kessinger (Wood River Sand Co.) v. Standard Oil Co., 245 Ill. App. 376, 380-88 (1925) (denying plaintiff's claim for damages when his excavation business was shut down due to the defendant's negligence because excavation was illegal); Desmet v. Sublett, 225 P.2d 141, 141-42 (N.M. 1950) (denying plaintiff's claim for compensation for lost profits from illegal operation of his truck, which had been wrongfully detained by the defendant); Harper v. Grasser, 150 P. 1175, 1176 (Wash. 1915) (finding that defendants were not entitled to damages on a counterclaim for being prevented by the plaintiff from fishing with a drag seine, given that this fishing method was prohibited).

250. [1986] 1 All E.R. 676 (Q.B.).

251. *Id.* at 679.

252. *Id.* at 689.

253. *See, e.g.*, McNichols v. J. R. Simplot Co., 262 P.2d 1012, 1015-16 (Idaho 1953) (determining that if illegal business is conducted in a certain building that is affected by a nearby nuisance, the owner will not be compensated for lost profits; however, he may recover

Deprivation of criminal profit may also be justified, at least in part, by the notion of retributive justice. If tort law enabled the offender to obtain the benefits of her crime, then criminal law would be undermined: the expected profit would offset the expected sanction and allow crime to pay. By preventing the offender from acquiring the criminal profit, criminal justice is preserved. Since criminal justice is based on retribution and deterrence, denial of liability also serves these twin goals.[254] From a retributive perspective denial of recovery for the loss of criminal profits sustains the necessary correlativity between crime and punishment.

In several tort cases, it was held that liability for crime-related losses should be denied if there was *statutory intent* to the effect that a person in breach of the relevant penal provision would be deprived of a private right of action.[255] The statutory-intent test has been used by both English and Australian courts: in England as a guideline for the application of the *ex turpi causa* defense,[256] and in Australia as a policy factor that may negate the duty of care.[257] This conceptual difference is not very significant from a theoretical standpoint. What matters is that illegality bars recovery whenever this is consistent with the statutory intent. Since criminal statutes are generally intended to advance the goals of retribution and deterrence, the question becomes whether the additional sanction of no-recovery in tort also serves these goals. Logically, whenever a court finds that denying liability is consistent with the statutory intent, it is actually finding that this denial serves retribution, deterrence, or both. Consequently, the application of *ex turpi causa*, or equivalent doctrinal frameworks, in light of the statutory intent, may be viewed, at least in part, as another manifestation of retributive justice.

---

for injuries to the structure itself); THE ILLEGALITY DEFENCE IN TORTS, *supra* note 194, at 40 n.181 ("The illegality defense may operate to affect a head of damage, rather than the entire claim").

254. THE ILLEGALITY DEFENCE IN TORTS, *supra* note 194, at 74.

> For the law to proscribe particular conduct and yet allow the offender successfully to sue to ensure that he or she profits from that conduct is, we believe, an unacceptable situation, and one that would, in effect, undermine the law. It would have the result that to allow someone to benefit from his or her wrongdoing would mean that "crime would pay."

Id.

255. *See* cases cited *infra* notes 256-57.

256. See, for example, *Cakebread v. Hopping Brothers Ltd.*, [1947] 1 K.B. 641, 654, in which the plaintiff employee claimed for injuries suffered as a result of the employer's breach of statutory and regulatory duties. *Id.* at 642. The employer raised the illegality defense, claiming that the plaintiff had aided and abetted the illegality. *Id.* The defense failed because, "[t]he policy of the . . . Act makes it plain that such a defence as that put forward here would be inconsistent with the intention of Parliament." *Id.* at 654. *See also* THE ILLEGALITY DEFENCE IN TORTS, *supra* note 194, at 79 ("[T]he illegality doctrine should operate in a way which supports the purpose of the rule which makes the conduct illegal in the first place. Conversely, if allowing the claim would not in any way undermine the purpose of the rule which made the conduct illegal, this rationale will not support the application of the doctrine.").

257. Henwood v. Mun. Tramways Trust (1938) 60 C.L.R. 438, 460-61 (Austl.).

The need to prevent the offender from being compensated for criminal penalty and from acquiring the benefits of her wrong, may, in fact, be deemed to derive from the more general desire to implement the intent of criminal statutes. It is always a part of the purpose of a penal provision that a violator will not be able to evade the prescribed sanction or obtain a benefit from such violation;[258] otherwise, the conduct would not be banned and penalized. It may be that these two derivatives do not exhaust the scope of statutory intent. However, the desire to implement the statutory intent, and its two alleged derivatives, are ultimately the by-products of the more abstract idea of preserving criminal justice.

d. Disallowing Liability Between Joint Offenders

A limited variation of the *ex turpi causa* defense provides that a party who consents to, and participates in, an immoral or illegal act cannot recover damages *from other participants* for the consequences of that act.[259] This formula has been consistently adhered to by the Supreme Court of Virginia.[260] For example, in *Zysk v. Zysk*,[261] the court rejected a woman's claim against her husband for a sexually transmitted disease contracted through sexual intercourse before marriage, given that pre-marital intercourse constituted the punishable misdemeanor of fornication.[262] In *Miller v. Bennett*,[263] the same court denied recovery for the injuries and subsequent death of a woman resulting from an illegal, negligently performed abortion.[264]

---

258. THE ILLEGALITY DEFENCE IN TORTS, *supra* note 194, at 79-80 (arguing that statutory-intent theory may explain denial of compensation for criminal punishment and its consequences). *But see* Weinrib, *supra* note 202, at 31. Weinrib opines that it is impossible to determine the legislative intent for "a problem regarding which the legislature did not express, and probably did not have, an intention." *Id.* In my view, any criminal provision has an implicit intent to maintain the penalty it expressly imposes.

259. This formulation seems broader than the maxim *in pari delicto potior est conditio defendentis*. The latter bars recovery only if the plaintiff and the defendant are "equally criminal." Wash. Gas Light Co. v. D.C., 161 U.S. 316, 327 (1895). *Ex turpi causa* does not require equality.

260. Lee v. Nationwide Mut. Ins. Co., 497 S.E.2d 328, 329 (Va. 1998); Zysk v. Zysk, 404 S.E.2d 721, 722 (Va. 1990); Miller v. Bennett, 56 S.E.2d 218, 218 (Va. 1949). Other courts and scholars have also recognized this principle. Gilmore v. Fuller, 65 N.E. 84, 86-87 (Ill. 1902); COOLEY, *supra* note 197, at 151.

261. 404 S.E.2d at 721.

262. *Id.* at 722.

263. 56 S.E.2d at 217.

264. *Id.* at 218-19; *accord* Hunter v. Wheate, 289 F. 604, 606-07 (D.C. Cir. 1923) (finding that a physician was not liable for injuries ensuing from an illegal and negligently-performed abortion); Nash v. Meyer, 31 P.2d 273, 274 (Idaho 1934); Castronovo v. Murawsky, 120 N.E.2d 871, 875 (Ill. App. Ct. 1954); Szadiwicz v. Cantor, 154 N.E. 251, 252 (Mass. 1926); Reno v. D'Javid, 369 N.E.2d 766, 767 (N.Y. 1977); Henrie v. Griffith, 395 P.2d 809 (Okla. 1964); Martin v. Morris, 425 S.W.2d 207, 207 (Tenn. 1931); Androws v. Coulter, 1 P.2d 320, 321 (Wash. 1931). For a criticism of this line of cases, see Hollister, *supra* note 194. *Cf.*

The principle of no-liability between participants in a criminal venture, at least where the crime committed was particularly risky, seems to have been recognized by English courts as well.[265] A similar principle, within the duty-of-care framework, was accepted and applied in Australia. In *Smith v. Jenkins*,[266] for example, the plaintiff and the defendant assaulted a car owner and fled with the car.[267] The plaintiff, who was a co-conspirator, sat in the passenger seat, and was injured by the negligent driving of the defendant.[268] The High Court of Australia denied recovery; Chief Justice Barwick found that no duty of care arises between joint participants in criminal conduct with regard to injuries or losses linked with such conduct.[269] Comparable conclusions were reached by his colleagues.[270] In subsequent cases the applicability of this principle was limited to cases of especially dangerous crimes.[271] In Scotland, it was similarly found that no duty of care is owed between persons engaged in a common criminal enterprise.[272] Here too, the rule was subsequently limited to cases of serious criminal activity.[273]

Several explanations may be given for the principle of no-recovery between participants in a criminal venture, with regard to harm or injury ensuing from the manner in which they performed the crime. However, most of these explanations can justify exclusion of liability only in certain types of criminal joint ventures. More important, most explanations seem to be unrelated to the notion of retributive justice. One possible explanation is that in certain cases courts are unable to determine the level of care owed by one criminal to her partner during the performance of their crime.[274] However, this explanation is

---

Goldnamer v. O'Brien, 335 S.W. 831, 832 (Ky. Ct. App. 1896) (finding no cause of action against persons who induced plaintiff to undergo an illegal abortion which resulted in personal injury); Bowlan v. Lunsford, 54 P.2d 666, 667-68 (Okla. 1936). *But see* True v. Older, 34 N.W.2d 700, 701 (Minn. 1948) (finding that if a pregnant woman is forced to submit to abortion by threats of such a serious nature as to destroy her own volition, there is no wrong on her part that bars recovery).

265. *See, e.g.*, Pitts v. Hunt, [1990] 3 All E.R. 344, 358 (Balcombe L.J., concurring); Colburn v. Patmore, 149 Eng. Rep. 999, 1003 ("[A] person who is declared by the law to be guilty of a crime cannot be allowed to recover damages against another who has participated in its commission.").

266. (1970) 119 C.L.R. 397 (Austl.).

267. *Id.* at 405.

268. *Id.*

269. *Id.* at 400.

270. *Id.* at 403 (Kitto J., concurring), 422 (Windeyer J., concurring), 426 (Owen J., concurring), 429, 433 (Walsh J., concurring); *accord* Progress & Properties Ltd. v. Craft, (1976) 135 C.L.R. 651, 668 (Austl.) ("A joint illegal activity may absolve the one party from the duty towards the other to perform the activity with care for the safety of that other.").

271. Jackson v. Harrison, (1978) 138 C.L.R. 438, 455-56 (Austl.).

272. The leading authority is *Lindsay v. Poole*, 1984 S.L.T. 269. It was followed, for example, in *Wilson v. Price*, 1989 S.L.T. 484 and *Sloan v. Triplett*, 1985 S.L.T. 294. *See also* Lord Hunter's dictum in *Winnik v. Dick*, 1984 S.L.T. 185, 189.

273. *See* Taylor v. Leslie, 1998 S.L.T. 1248; Weir v. Wyper, 1992 S.L.T. 579, 581-82.

274. *See, e.g.*, *Jackson* 138 C.L.R. at 455-56.

not applicable to all cases of criminal joint ventures. It is relevant only where the crime itself is of such a risky nature that it is impossible to say when the risk created by one of the participants becomes "unreasonable."[275] This explanation conforms to the corrective structure of tort law.

A closely related explanation is that criminal conduct, by its very nature, defies legal norms; therefore, one who associates with another to commit a crime cannot expect the other to comply with legal standards while perpetrating the crime. This supposition is implicit in the agreement to commit the crime. Disallowing the participants to sue each other for resulting injuries, or losses, may be roughly described as a variation of the principle of *volenti non fit injuria*.[276] This explanation seems to apply only to risky crimes, not to all crimes, because a person who agrees to carry out a non-dangerous crime with another does not expect the other to create unreasonable risks during its commission.[277] As denial of recovery under this explanation is founded on the concept of implied consent within the bilateral interaction, it also conforms to corrective justice.

A third possible account for the joint-offenders version of *ex turpi causa* is that resolving disputes between criminal offenders, concerning the consequences of their joint crime, may impair the dignity and reputation of the courts and the legal system (at least in cases of serious or heinous illegality). The court may seem to operate as an "underworld peacemaker."[278] This is obviously an institutional concern, having nothing to do with retribution.

A more secure foundation for denying relief, though more limited in its application—and for that reason fairer in its operation—is to say that the plaintiff must fail when the character of the enterprise in which the parties are engaged is such that it is impossible for the court to determine the standard of care which is appropriate to be observed.

Id. (*cited with approval in* Pitts v. Hunt, [1990] 3 All E.R. 344, 358).

275. This may be so, for example, where the plaintiff and the defendant participate in a reckless joy ride. If courts were totally incapable of setting the standard of care owed by criminals to their partners during the performance of their crime, this would be a good reason to disallow tort claims between joint offenders altogether. But this does not seem to be the case.

276. Miller v. Bennett, 56 S.E.2d 217, 219 (Va. 1949) ("[C]onsent or participation in an immoral or unlawful act by plaintiff precludes recovery for injuries sustained as a result of that act, on the maxim volenti non fit injuria."); Smith v. Jenkins, 119 C.L.R. 397, 422 (Windeyer J., concurring) (observing that disallowing liability between joint offenders may be regarded as an extension of the *volenti* defense); Weinrib, *supra* note 202, at 34-35 (discussing the possible elision of *ex turpi causa* in the *volenti* defense). *Cf.* Gala v. Preston, 172 C.L.R. 243, 254 (Austl. 1991) ("[T]he participants could not have had any reasonable basis for expecting that a driver of the vehicle would drive it according to ordinary standards of competence and care.").

277. *But cf.* COOLEY, *supra* note 197, at 152 n.3:

> [I]t may be said that usually it is only reckless parties who plan and engage in unlawful action, and therefore the want of care and prudence on the part of the associates ought to be assumed as one of the probable concomitants, the risks of which each must be understood to take upon himself when he engages in the unlawful act.

Cooley emphasized the nature of criminals rather than the nature of the specific crime.

278. As observed by the British Law Commission in a similar context, "[r]ather than stooping to the indignity of inquiring into the relative merits and demerits of the parties . . . courts should simply leave matters as they are." THE EFFECT OF ILLEGALITY ON CONTRACTS,

There are at least two possible explanations that link (albeit rather loosely) the unique version of *ex turpi causa* discussed here with the notion of retributive justice. First, under criminal law, when several persons knowingly contribute to the joint commission of a crime, their acts are non-severable for the purposes of legal responsibility. In other words, if a person is guilty of a crime she is responsible for all of its constituents, including those performed by her accomplices.[279] Thus, it may be said that, in the current context, the plaintiff is legally responsible for the act that resulted in her own harm or injury. Therefore, she cannot claim damages from other participants for that harm or injury.[280] It follows that the joint-offenders version of *ex turpi causa* supports a general principle of criminal responsibility and thereby advances the goals of criminal law.

Second, one could argue that a criminal who brings suit against her accomplice, with regard to the upshot of their criminal joint venture, often attempts to retrieve the benefits of the criminal agreement and conduct. Applying *ex turpi causa* deprives the plaintiff of these benefits, and thereby reinforces criminal justice and its underlying goals. There are two versions of this argument. According to the narrow version, denying recovery is justified only when the plaintiff attempts to recoup the expected proceeds of the respective crime. This version is actually independent of the fact that the plaintiff and the defendant were involved in a joint venture. Allowing a criminal to recover the proceeds of her crime from any person, accomplice or other, would undermine the criminal punishment, as observed in the previous subsection.[281]

According to the broad version, denying recovery is justified whenever it prevents enforcement of an agreement to commit a crime or any part of it. If two persons agree to commit a crime together, they have certain expectations that are incorporated, either explicitly or implicitly, into their initial agreement. Among other things, a participant may expect her partner to perform his part with prudence. Allowing her to claim damages for an injury caused by his negligence during their joint action, would put into effect their illegal

---

supra note 202, at 87. *Contra Smith*, 119 C.L.R. at 432 (Walsh J., concurring) ("However, I do not think that the essential reason for a rule by which the courts refuse to recognize a right of action in some cases of criminality is not a shrinking by the court from the seamy facts of life or a scrupulous regard for its dignity and reputation.")

279. MODEL PENAL CODE § 2.06(2)(c) (1962) ("A person is legally accountable for the conduct of another person when . . . he is an accomplice of such other person in the commission of the offense."). A person may be regarded as an accomplice of another person in the commission of an offense if "with the purpose of promoting or facilitating the commission of the offense," he aids or agrees or attempts to aid such other person in planning or committing it. *Id.* § 2.06(3)(a)(ii); Am. Motorcycle Ass'n v. Superior Court, 578 P.2d 899, 904 (Cal. 1978) ("[A]ll members of a 'conspiracy' or partnership [to commit a crime] are equally responsible for the acts of each member in furtherance of such conspiracy."); Gilmore v. Fuller, 65 N.E. 84, 87 (Ill. 1902).

280. *Gilmore*, 65 N.E. at 87-88; *Smith* 119 C.L.R. at 404 (Kitto J., concurring).

281. *See supra* Subsection IV.B.1.c.

agreement. For example, in cases of negligently performed illegal abortions, one could say that allowing the woman to claim damages is equivalent to enforcing an implied term concerning the abortionist's level of care. By invalidating an agreement to commit a crime, whether or not it can be regarded as a "contract" in the legal sense, the court deprives the offender of any advantage accruing from the agreement and thereby upholds the severity of the criminal sanction imposed on each of the contracting parties. This explanation seems to complement the implied-consent argument discussed above: if the crime is so risky that its perpetrator cannot expect her accomplice to comply with legal standards during its commission, the latter applies; whereas if it can be implied that each participant expected the other to act with due care the former applies.

2. The Innocence Requirement in Malpractice Claims Against Defense Attorneys

In ordinary malpractice proceedings the plaintiff must prove four elements to establish her cause of action: duty of care, breach of that duty, damage, and a causal link between the breach and the damage.[282] Yet, when a criminal defense attorney is sued for not achieving a less onerous outcome for her client in criminal proceedings, a fifth element is added in most jurisdictions. In some, the plaintiff has to prove that she is *factually* innocent of the underlying charge.[283] In other jurisdictions, the plaintiff must obtain post-conviction relief as a precondition to recovery;[284] stated differently, she must prove that she is *legally* innocent under the rules of criminal law and procedure. It may be generalized that in most jurisdictions a convicted felon who files a suit against her defense attorney for malpractice is required to show that she is innocent, either factually or legally. Although this requirement is somewhat related to one of the modern applications of *ex turpi causa*,[285] there is a notable difference between the two. *Ex turpi causa* is a defense that operates only when raised and

---

282. David H. Potel, Comment, *Criminal Malpractice: Threshold Barriers to Recovery against Negligent Criminal Counsel*, 1981 DUKE L.J. 542, 543-45 (1981).

283. Levine v. Kling, 123 F.3d 580, 582 (7th Cir. 1997); Wiley v. County of San Diego, 966 P.2d 983, 985 (Cal. 1998); Moore v. Owens, 698 N.E.2d 707, 709 (Ill. App. Ct. 1998); Ray v. Stone, 952 S.W.2d 220, 224 (Ky. Ct. App. 1997); Glenn v. Aiken, 569 N.E.2d 783, 785 (Mass. 1991); Labovitz v. Feinberg, 713 N.E.2d 379, 382 (Mass. App. Ct. 1999); Rodriguez v. Nielsen, 609 N.W.2d 368, 374 (Neb. 2000); Mahoney v. Shaheen, Cappiello, Stein & Gordon, P.A., 727 A.2d 996, 998-99 (N.H. 1999); Carmel v. Lunney, 511 N.E.2d 1126, 1128 (N.Y. 1987); Biegen v. Paul K. Rooney, P.C., 703 N.Y.S.2d 121, 121-22 (N.Y. App. Div. 2000); Bailey v. Tucker, 621 A.2d 108, 113 (Pa. 1993); Brown v. Theos, 526 S.E.2d 232, 235 (S.C. Ct. App. 1999).

284. Shaw v. State, 816 P.2d 1358, 1360 (Alaska 1991); Steele v. Kehoe, 747 So. 2d 931, 933 (Fla. 1999); Morgano v. Smith, 879 P.2d 735, 737 (Nev. 1994); Stevens v. Bispham, 851 P.2d 556, 561 (Or. 1993); Gibson v. Trant, 58 S.W.3d 103, 107 (Tenn. 2001); Peeler v. Hughs & Luce, 909 S.W.2d 494, 495, 497-98 (Tex. 1995); Adkins v. Dixon, 482 S.E.2d 797, 801 (Va. 1997).

285. *See supra* Subsection IV.B.1.c.

substantiated by the defendant, whereas the innocence requirement adds an essential element to the plaintiff's cause of action.[286]

Three justifications are suggested for the innocence requirement. First, it is said that the additional element ensures the readiness of lawyers to take criminal defense cases, especially when the defendant is indigent and there is no financial motivation to do so.[287] However, encouraging lawyers to represent the *needy* criminal-defendant does not, by itself, justify a general innocence requirement. It cannot justify the application of this requirement where the motivation to represent could exist without it. Moreover, if the doctrine is intended to encourage criminal lawyers to accept indigent clients, it seems peculiar that the incentive depends on the fortuitous fact of the client's innocence. Oddly, the innocence requirement only creates an incentive to represent the guilty.

The second justification is that the innocence requirement serves as a deterrent to frivolous litigation.[288] It is sometimes said that inmates may fill their free time pursuing civil actions against their former attorneys;[289] however, statistics show that prisoners do not have such a litigious nature.[290] Furthermore, the fear of frivolous claims should be dealt with through the laws of evidence and procedure, rather than through substantive law.

The third, and most plausible justification, is that the innocence requirement serves as a means for preserving criminal justice by ensuring that a convicted felon is not compensated for the criminal punishment and its consequences unless she did not deserve to be punished in the first place. The innocence requirement thereby prevents stultification of the criminal punishment and advances the goals of criminal law.[291]

The third justification has attracted several criticisms. For example, it has been argued that "courts that manipulate tort remedies to further criminal law goals risk reckless, and unnecessary doctrinal blurring."[292] However, at this stage the reader might not find this "doctrinal blurring" so disturbing; the same "blurring" has existed for centuries in the areas of punitive damages (discussed

---

286. However, in at least one jurisdiction that has adopted the *legal* innocence requirement, the defendant can raise the issue of the plaintiff's *factual* guilt as an affirmative defense. Shaw v. State, 861 P.2d 566, 572 (Alaska 1993).

287. *Glenn*, 568 N.E.2d at 788 (noting that the innocence requirement encourages attorneys to take criminal defense cases); *Gibson*, 58 S.W.3d 103 at 115-16 (stating that the requirement of post-conviction relief provides an incentive to represent indigent clients).

288. Susan M. Treyz, Note, *Criminal Malpractice: Privilege of the Innocent Plaintiff?*, 59 FORDHAM L. REV. 719, 729-30 (1991).

289. *Stevens*, 851 P.2d at 563 (citing *Shaw*, 816 P.2d at 1361).

290. Note, *Recent Case: Criminal Defense -- Legal Malpractice -- Nevada Limits the Rights of Inmates to Sue Their Former Attorneys for Malpractice - Morgano v. Smith, Nos. 24958, 25215, 1994 WL 419906 (Nev. Aug. 10, 1994)*, 108 HARV. L. REV. 501, 503-04 (1994) [hereinafter *Recent Case*]. "[S]tatistics show that prisoners do not file criminal malpractice suits in large numbers." *Id.* at 503.

291. *Id.* at 501 (stating that "the force driving the recent decisions is a desire to promote the criminal law goals of deterring and punishing crime.").

292. *Id.* at 505.

below) and *ex turpi causa* (examined above). In addition, what may seem, at first glance, to be "doctrinal blurring" may, in fact, guarantee consistency within the law. Returning to the words of Judge McLachlin, the law refuses to give by its right hand (tort law) what it takes away by its left hand (criminal law).[293]

Another criticism is that the innocence requirement fosters attorney negligence and thereby increases *expected* sanctions and undermines the retributive effort.[294] However, this argument misapprehends the idea of retributive justice. Unlike deterrence, retributive justice focuses on the fairness of the *actual* punishment, in light of the circumstances of the specific case, not on the *expected* sanction, which combines the severity of the penalty with the probability of crime-detection, indictment, and conviction. The fact that professional negligence may increase the probability of sentencing is irrelevant from a retributive perspective. What matters is that the severity of the actual sanction is proportional to the gravity of the crime.

A better formulated criticism would be that the innocence requirement may lead to an increase in the severity of criminal punishment, as lawyers would not exert their best efforts to acquit their clients or minimize their punishments. This fear seems exaggerated. The innocence requirement does not affect the incentive not to be negligent in representing innocent clients because these clients are entitled to compensation in cases of malpractice.[295] The innocence requirement may reduce the lawyer's incentive to make an effort for the guilty. But if a client is convicted and sentenced for a crime she actually committed, it can rarely be said that the attorney's negligence resulted in an undue punishment. Perhaps the attorney could have obtained a lesser sanction in such situations, but this does not make the actual punishment excessive in the retributive sense.

V. THE ROLE OF RETRIBUTIVE JUSTICE IN EXPANDING LIABILITY

A. Punishing the Injurer for Illicit Conduct

1. Punitive Damages

Punitive damages have been recognized in the common law of torts for centuries, and are now firmly entrenched in most Anglo-American jurisdictions.[296] Clearly, a doctrine that awards extra-compensatory damages is

293. Hall v. Herbert, [1993] 101 D.L.R. (4th) 129, 160.

294. *Recent Case*, *supra* note 290, at 505-06.

295. There is, perhaps, only a marginal effect, because innocent clients will not always be able to prove their innocence in a civil action.

296. The first reported American punitive damages case is Genay v. Norris, 1 S.C.L. (1 Bay) 6, 8 (1784); *accord* Denver & Rio Grande Ry. v. Harris, 122 U.S. 597, 609 (1887); Day v. Woodworth, 54 U.S. 363, 371 (1851); Lamb v. Cotogno, (1987) 146 C.L.R. 1 (Austl.); Kuddus v. Chief Constable of Leicestershire Constabulary, [2001] UKHL 29, [2001] 2 W.L.R. 1789 (appeal taken from Eng.); Tullidge v. Wade, (1769) 95 Eng. Rep. 909 (K.B.); Wilkes v. Wood,

inconsistent with theories of corrective justice that focus on rectification of the harm caused.[297] One might ask what, then, is the purpose of punitive damages? In the past, it was frequently said that the aim of these awards was "to punish and deter." This language has been used in court decisions,[298] professional and academic literature,[299] and even jury instructions.[300] However, this phrase is somewhat misleading, because punishment, as the word is currently understood, is not a purpose but a mechanism. Punishment may be defined as "imposing a sanction,"[301] and it may have various purposes, such as retribution, deterrence, appeasement of the victim, incapacitation of the wrongdoer, and education. Saying that punitive damages are meant to punish is consequently tautological. It may be transformed into the odd sentence, "the purpose of imposing this sanction is to impose a sanction."

Still, it is clear that the word "punish" as used in the phrase "to punish and deter," has not been used in vain. The truth is that courts and scholars sometimes use the terms "punishment" and "retribution" interchangeably. While such usage is perplexing it is still accepted. Therefore, it seems that when courts say that punitive damages are supposed to "punish and deter," they suggest that punitive damages are aimed at retribution and deterrence. This is the most reasonable explanation for the classical punish-deter dichotomy.[302] In more recent American decisions, the terminology has become more accurate.

(1763) 98 Eng. Rep. 489, 490 (K.B.); Huckle v. Money, (1763) 95 Eng. Rep. 768, 768-69 (K.B.) (introducing the term "exemplary damages" to explain an award that exceeded actual damage). The idea of non-compensatory damages was recognized in ancient legal systems as well. *See* Michael Rustad & Thomas Koenig, *The Historical Continuity of Punitive Damages Awards: Reforming The Tort Reformers*, 42 AM. U. L. REV. 1269, 1285-86 (1993) (discussing the "ancient lineage" of concepts like punitive damages by reference to Babylonian Hammurabi Code, the Hindu Code of Manu, the Old Testament, the New Testament, and Roman Law).

297. Weinrib, *supra* note 7, at 135-36 n.25; *see also* Murphy v. Hobbs, 5 P. 119, 125 (Colo. 1884) (stating that punishment and compensation should be kept "separate and distinct"); 2 SIMON GREENLEAF, A TREATISE ON THE LAW OF EVIDENCE 240 n.2 (16th ed. 1899) (positing that punitive damages are alien to compensatory tort law).

298. BMW of N. Am., Inc. v. Gore, 517 U.S. 559, 568 (1996); Newport v. Fact Concerts, Inc., 453 U.S. 247, 266-67 (1981); Hodges v. S.C. Toof & Co., 833 S.W.2d 896, 900 (Tenn. 1992).

299. RESTATEMENT (SECOND) OF TORTS § 908 cmt. a (1979); Howard A. Denemark, *Seeking Greater Fairness When Awarding Multiple Plaintiffs Punitive Damages for a Single Act by a Defendant*, 63 OHIO ST. L.J. 931, 935-36 (2002).

300. RONALD W. EADES, JURY INSTRUCTIONS ON DAMAGES IN TORT ACTIONS 98 (3d ed. 1993) (quoting standard jury instruction according to which the purposes of punitive damages is to punish and deter).

301. *See* George P. Fletcher, *The Place of Victims in the Theory of Retribution*, 3 BUFF. CRIM. L. REV. 51, 54 (1999).

302. Ellis, *supra* note 22, at 3-4 (stating that punitive damages serve deterrence and retribution); David G. Owen, *A Punitive Damages Overview: Functions, Problems and Reform*, 39 VILL. L. REV. 363, 373-74 (1994); Catherine M. Sharkey, *Punitive Damages as Societal Damages*, 113 YALE L.J. 347, 356-57 (2003); Sunstein, et al., *supra* note 54, at 2081; Malcolm E. Wheeler, *The Constitutional Case for Reforming Punitive Damages Procedures*, 69 VA. L. REV. 269, 310 (1983).

For example, in *State Farm Mutual Automobile Insurance Co. v. Campbell*[303] the Supreme Court of the United States stated that "[c]ompensatory damages 'are intended to redress the concrete loss that the plaintiff has suffered by reason of the defendant's wrongful conduct.'. . . By contrast, punitive damages serve a broader function; they are aimed at deterrence and retribution."[304]

Between deterrence and retribution, retribution seems to be the dominant goal. Proponents of the economic analysis of the law offer a deterrence-based rationale for this doctrine. Their main argument is that punitive damages may be used to overcome problems of under-enforcement.[305] Under this perception, total damages should equal the harm, multiplied by the reciprocal of the probability that the defendant will be found liable in cases where she should be found liable; punitive damages would then consist of the excess of total damages over compensatory damages.[306] However, this theory has very weak descriptive power and severe normative deficiency. From a descriptive standpoint, an under-enforcement rationale can hardly explain why punitive damages are unusual in all common law jurisdictions and why they hinge on the defendant's state of mind.[307] From a normative perspective, administrative and criminal law may also be regarded as efforts to compensate for private under-enforcement of law; if tort law also attempts to handle under-enforcement through punitive damages, over-deterrence may ensue.[308]

Thus, retribution is the better explanation for punitive damages. That is so not only on the theoretical level but also in practice. Empirical evidence suggests that juries do not attempt to promote optimal deterrence, but to "punish" wrongdoing, with at most a signal designed to ensure that certain misconduct will not happen again.[309] Ordinary people do not spontaneously think in terms of optimal deterrence when asked questions about appropriate punishment, and it is very hard to get them to think in these terms.[310] Jurors assume their roles with retributive intuitions, and "it remains unclear whether and to what extent the courtroom can overcome those intuitions."[311]

As previously mentioned, punitive damages are accepted in nearly all common law jurisdictions; however, they are highly unusual. Awarding non-compensatory damages is inconsistent with the corrective structure of tort law. Therefore, punitive damages are merely awarded at the margins, within the

303. 538 U.S. 408 (2003).

304. *Id.* at 416 (citations omitted). *See also id.* at 417 (opining that punitive damages awards "serve the same purposes as criminal penalties"); Pac. Mut. Life Ins. Co. v. Haslip, 499 U.S. 1, 19 (1991) (discussing punitive damages close relationship with criminal penalties).

305. WILLIAM M. LANDES & RICHARD A. POSNER, THE ECONOMIC STRUCTURE OF TORT LAW 160-65, 184-85, 223-24 (1987); SHAVELL, *supra* note 10, at 148; A. Mitchell Polinsky & Steven Shavell, *Punitive Damages: An Economic Analysis*, 111 HARV. L. REV. 869, 873-74 (1998); Sharkey, *supra* note 302, at 365-68; Sunstein, et al., *supra* note 54, at 2075, 2082.

306. For further discussion of this rationale see sources cited *supra* note 305.

307. Zipursky, *supra* note 115, at 95.

308. Sunstein, et al., *supra* note 54, at 2084.

309. *Id.* at 2085.

310. *Id.* at 2111.

311. *Id.*

prevention of abominable disproportion paradigm. They are used only when the courts determine that a compensatory award is an *extremely lenient* sanction with regard to the gravity of the defendant's conduct. In other cases, the discrepancy between the gravity of the wrong and the severity of the sanction is left untouched for the sake of corrective justice. According to the Supreme Court, it should be presumed a plaintiff has been made whole for his or her injuries by compensatory damages; so "punitive damages should only be awarded if the defendant's culpability, after having paid compensatory damages, is so reprehensible as to warrant the imposition of further sanctions to achieve retribution or deterrence."[312]

Because retribution is the primary goal of punitive damages, courts have been cautious not to award such damages in violation of the principles of retributive justice. According to the principle of cardinal proportionality a sanction should not be too harsh or too lenient with regard to the absolute gravity of the wrong committed. In particular, punitive damages should not increase the *overall burden* imposed on the defendant beyond what is due.[313] Consequently, the Supreme Court has determined that an award of punitive damages must reflect the gravity of the respective wrong,[314] and that the "Due Process Clause of the Fourteenth Amendment prohibits the imposition of grossly excessive or arbitrary punishments on a tortfeasor."[315] A jury's punitive damages award is, therefore, subject to substantive due process review under the Fourteenth Amendment.

In *BMW v. Gore*, the Supreme Court counseled that in reviewing whether punitive damages awards are grossly excessive under the Due Process Clause courts ought to consider three guideposts: (1) the degree of reprehensibility of the defendant's conduct; (2) the disparity between the actual or potential harm suffered by the plaintiff and the punitive damages award; and (3) the difference

312. State Farm Mut. Auto. Ins. Co. v. Campbell, 538 U.S. 408, 419 (2003) (citing BMW of N. Am. v. Gore, 517 U.S. 559, 575 (1996) ("[T]he most important indicium of the reasonableness of a punitive damages award is the degree of the reprehensibility of the defendant's conduct.")).

313. Overall burden consists of sanctions of a legal nature, such as criminal, administrative, and civil as well as sanctions of an extra-legal nature, such as reputational harm.

314. Day v. Woodworth, 54 U.S. 363, 371 (1851). In *Day*, the Court also stated that "the damages assessed depend on the circumstances, showing the degree of moral turpitude or atrocity of the defendant's conduct. . . ." *Id.*

315. *Campbell*, 538 U.S. at 416; *accord* Cooper Indus., Inc. v. Leatherman Tool Group, Inc., 532 U.S. 424, 434 (2001) (stating that a penalty violates the Due Process Clause if it is "grossly disproportional to the gravity of . . . defendants' offenses.") (quoting U.S. v. Bajakajian, 524 U.S. 321, 334))). The Fourteenth Amendment prohibits imposition of excessive penalties by the States in two ways. *Cooper*, 532 U.S. at 434. First, it makes the Eighth Amendment's prohibition against excessive fines and cruel and unusual punishments applicable to the States. Gideon v. Wainwright, 372 U.S. 335, 341-42 (1963). Second, the Due Process Clause, through its own power, prohibits the States from imposing "grossly excessive" punishments. *BMW*, 517 U.S. at 562; *accord* TXO Prod. Corp. v. Alliance Res. Corp., 509 U.S. 443, 453-54 (1993).

between the punitive damages awarded and the civil penalties authorized or imposed in comparable cases.[316]

The first, and most important, guidepost[317] can be easily explained in retributive terms. Determining the severity of the sanction in light of the degree of reprehensibility is clearly an exercise of retributive justice. The third guidepost seems to be an attempt to maintain ordinal proportionality: comparable wrongs deserve similar sanctions. The second guidepost can also be linked with the notion of retributive justice because the potential harm is one of the primary indicators of the gravity of the defendant's conduct. However, it seems that the second guidepost is also intended to prevent the plaintiff from acquiring an unreasonable windfall, and not just to ensure that the defendant's sanction is due.

In any event, the first guidepost is dominant. True, the Supreme Court found in *Campbell* that a single-digit ratio between punitive and compensatory damages is more likely to accord with due process than awards with ratios in range of 500 to 1 (or even 145 to 1);[318] however, the Court emphasized that greater ratios may be consistent with the Due Process Clause "where 'a particularly egregious act has resulted in only a small amount of economic damages.'"[319] Egregious conduct can give rise to various types of sanctions, legal (criminal, administrative, civil, or disciplinary), and extra-legal (such as reputational harm). Thus, to avoid disproportion between the overall burden imposed on the defendant and the gravity of her wrong, these sanctions must be considered when deciding whether punitive damages may be awarded and in determining their amount. In many jurisdictions, this consideration is kept in mind.[320]

Another feature of the law of punitive damages which can also be explained in retributive terms is the common law's consistent and express refusal to award punitive damages against municipalities.[321] The Supreme Court found that absolute immunity from punitive damages in common law extends to claims under 42 U.S.C. § 1983.[322] Courts distinguish between compensation for injuries inflicted by a municipality's officers or agents and

316. *BMW*, 517 U.S. at 575.

317. *Id.* at 575 ("[T]he most important indicium of the reasonableness of a punitive damages award is the degree of reprehensibility of the defendant's conduct.").

318. 538 U.S. at 425.

319. *Id.* Perhaps *TXO* represents such an exceptional case because while compensatory damages were $19,000, the punitive damages were set at $10M after the Court determined that the defendant had pursued "a malicious and fraudulent course." *TXO Prod. Corp.*, 509 U.S. at 462. (1993).

320. Saunders v. Gilbert, 72 S.E. 610, 615 (N.C. 1911) (finding that pecuniary punishment in a criminal procedure can be considered in reduction of punitive damages); Annotation, *Assault: Criminal Liability as Barring or Mitigating Recovery of Punitive Damages*, 98 A.L.R.3d 870 (1980); Weinrib, *supra* note 202, at 46.

321. Newport v. Facts Concerts, Inc., 453 U.S. 247, 260. *See also* Joel E. Smith, Annotation, *Recovery of Exemplary or Punitive Damages from Municipal Corporation*, 1 A.L.R.4th 448 (1980).

322. *Newport*, 453 U.S. at 271.

the award of punitive damages for bad-faith conduct by the same officers or agents.[323] Compensation is an obligation properly shared by the municipality itself, whereas punishment is properly applied only to the actual wrongdoers.[324] The underlying rationale is that innocent citizens should be protected from unjust punishment.[325] An award of punitive damages against a municipality *only* punishes the taxpayers, either through increased municipal taxes or by a decrease in public services, although they took no part in the wrongdoing.[326] Under ordinary principles of retribution, it is the wrongdoer himself who is made to suffer for his unlawful conduct. Awarding punitive damages against a municipality violates this principle.

An interesting question concerns judicial readiness to award punitive damages against private firms. Firms are not persons, and when punitive damages are imposed on them those who suffer may not be wrongdoers at all. A punitive damages award may end up injuring not the firm so much as dispersed stockholders (if the firm absorbs the cost or a fraction of it), consumers (if the firm spreads the cost *ex post* through price increases), and employees (if the firm is forced to perform cutbacks, or goes into liquidation), who were not involved in the wrongful act. It is far from clear that juries awarding punitive damages are aware of this point, and it is also far from clear that they can be easily convinced that this point is correct.[327]

A possible explanation for the difference in attitude toward municipalities as against firms is that the former are supposed to perform acts for the common good, while the latter exist for private gain.[328] A malicious act performed by an officer of a municipal authority cannot be said to have been carried out in the name of the municipality and in the furtherance of its interests. A firm, on the other hand, may advance its interests by willfully and maliciously harming others. A malicious act by a firm's agent in the course of her employment can be regarded as an act performed in pursuit of the firm's interests. The fact that punishing the firm may also harm innocent parties should not prevent such punishment. After all, innocent third parties are sometimes harmed even when an individual is punished.[329] From a retributive standpoint, the key question is whether or not the firm deserves to be punished. Admittedly, this explanation is debatable, so the municipality/private-corporation distinction in the law of punitive damages seems somewhat anomalous.

2. Relaxing Fundamental Preconditions of Tort Liability

A few tort doctrines that temper traditional and fundamental preconditions

---

323. *Id.* at 262-66.

324. *Id.* at 263.

325. *Id.*

326. *Id.* at 267; McGary v. President & Council of Lafayette, 12 Rob. 674, 677 (La. 1846).

327. Sunstein, et al., *supra* note 54, at 2114 n.157; *accord* Ellis, *supra* note 22, at 66-67; Kaplow & Shavell, *supra* note 6, at 1067-68.

328. Hunt v. Boonville, 65 Mo. 620, 625 (1877).

329. For example, if a person is incarcerated, her relatives may lose her financial support.

of tort liability (i.e. actual harm, causation in fact, and proximate causation) seem to have been developed under the influence of retributive considerations.

The first doctrine is the collateral source rule. This rule holds that the injurer's liability is not reduced when the plaintiff receives benefits from other sources, even if those benefits have partially or wholly mitigated her loss. Thus, the victim may recover for harm that she did not actually incur.[330] Although there may be a modern economic explanation for the collateral source rule,[331] the late Professor Fleming observed that a "constantly recurring refrain, with strong overtones of moral outrage, is that the defendant is a wrongdoer who should not be 'let off' from any portion of what is his due by the exertions and foresight of his victim or those who stood by him in his hour of need."[332] Accordingly, the collateral source rule "condones multiple recovery to avoid giving the tortfeasor a 'windfall.'"[333] Fleming indicated that this extreme position was based largely on obsolete prejudice regarding the function of tort law, later referred to as "punitive."[334] Another author remarked that "notions of retribution" could be found in the "rhetoric that surrounds the collateral source rule."[335] However, the presumed link between the collateral source rule and retribution is dubious. The scope of damages usually depends on the extent of the plaintiff's injury, not on the degree of the defendant's fault. Since the extent of the plaintiff's injury is usually fortuitous, it cannot be deemed an appropriate measure for the retributively just sanction, and cannot be adhered to so ardently on retributive grounds.[336]

A second doctrine that is sometimes said to have a retributive goal is an exception to the general rule that applies to proof of causation. The imposition of liability usually depends on the plaintiff's showing that "her injuries were caused by an act of the defendant or by an instrumentality under the defendant's control."[337] In many jurisdictions, an exception to this rule arises "[w]here the conduct of two or more actors is tortious, and it is proved that harm has been caused to the plaintiff by only one of them, but there is uncertainty as to which one caused it." Under the exception, each actor has the

---

330. Steven B. Hantler, et al., *Moving Toward the Fully Informed Jury*, 3 GEO. J. L. & PUB. POL'Y 21, 24-31 (2005) (discussing the collateral source rule); Catherine M. Sharkey, *Unintended Consequences of Medical Malpractice Damages Caps*, 80 N.Y.U. L. REV. 391, 457-58 (2005).

331. *See, e.g.*, SHAVELL, *supra* note 10, at 142-43.

332. Fleming, *supra* note 70, at 1483.

333. *Id.* at 1544.

334. *Id.*

335. Schwartz, *Liability Insurance*, *supra* note 62, at 327.

336. *Cf.* Note, *Unreason in the Law of Damages: The Collateral Source Rule*, 77 HARV. L. REV. 741, 749 (1964) ("Even the most flagrant wrongdoer ordinarily is not liable for damages unless harm in fact results from his conduct; and when harm does occur, his liability depends, not on the degree of fault, but on the extent of injury, the physical idiosyncrasies and earning capacity of the injured person, and on the latter's own degree of culpability.").

337. Sindell v. Abbott Labs., 607 P.2d 924, 928 (Cal. 1980); *accord* RESTATEMENT (SECOND) OF TORTS § 433B(1) (1965).

burden of proving that he did not cause that harm.[338] This is sometimes called "the alternative-liability theory."[339] The consequences of this rule may be quite inconsistent with corrective justice. Liability may be imposed on various persons who caused no harm to the plaintiff, just because they acted negligently. The reason for this exception is the injustice that would result from permitting proven wrongdoers, one or some of whom inflicted injury on an entirely innocent plaintiff, to escape liability simply "because the nature of their conduct and the resulting harm" somehow "made it difficult or impossible to prove which of them . . . caused the harm."[340] A reputable tort scholar observed that this argument "becomes more a penal [retributive] argument than a tort [corrective] argument."[341] However, the retributive effect is incomplete because the defendant can exonerate herself from liability for wrongful conduct by proving that her conduct was not the cause-in-fact of the plaintiff's injury.

A retributive rationale may also be attributed to a more liberal exception to the general proof-of-causation requirement, namely the doctrine of market-share liability, first recognized by the Supreme Court of California in the renowned DES case, *Sindell*.[342] Under this theory, a plaintiff who sustains an injury or illness by using a fungible product and cannot identify the specific manufacturer whose product caused her injury or illness, can sue the manufacturers of a *substantial* share of the relevant product at the time of consumption.[343] Each defendant will be held liable in an amount determined by its share of the market, unless it can prove that it could not have manufactured the product that actually caused plaintiff's harm.[344] Once again, a person may be held liable for harm that he did not cause, solely because his conduct was negligent. Moreover, since the plaintiff is not required to join all manufacturers of the relevant product, at the relevant time, it is possible that none of those who may be eventually found liable actually manufactured the specific product that harmed the plaintiff.[345] The doctrine of market-share liability is, therefore,

---

338. RESTATEMENT (SECOND) OF TORTS § 433B(3) (1965); *accord Sindell*, 607 P.2d at 928; Martin v. Abbott Labs., 609 P.2d 368, 375 (Wash. 1984); Collins v. Eli Lilly & Co., 342 N.W.2d 37, 45 (Wis. 1984).

339. Victor E. Schwartz & Liberty Mahshigian, *Failure to Identify the Defendant in Tort Law: Towards a Legislative Solution*, 73 CAL. L. REV. 941, 946-49 (1985) (discussing the alternative liability theory).

340. RESTATEMENT (SECOND) OF TORTS § 433B(3) cmt. f (1965).

341. Richard W. Wright, *Causation in Tort Law*, 73 CAL. L. REV. 1735, 1818 (1985).

342. 26 Cal.3d at 936. The doctrine was later adopted in other jurisdictions. *See, e.g.*, Hymowitz v. Eli Lilly & Co., 539 N.E.2d 1069, 1078 (N.Y. 1989). Other courts applied similar theories of liability in cases of the same type. *See, e.g., Martin*, 689 P.2d at 382 (finding that joinder of a substantial share of the market is not required and stating that "a particular defendant's potential liability is proportional to the probability that it caused plaintiff's injury"); *Collins*, 342 N.W.2d at 50, 53 (finding that joinder of a substantial share of the market is not required and then stating that damages should be apportioned on a relative-fault basis and not on a market-share basis).

343. *Sindell*, 607 P.2d at 937.

344. *Id.*

345. *Id.* at 936-37.

a more significant inroad on the corrective structure of tort law. It theoretically enables courts to impose tort liability on those who created certain risks, even though none of them *caused* the plaintiff's injury or illness.

Another principle (or better, tendency) that seems to rest on retributive grounds applies within the conceptual framework of proximate causation. The fact that a person's conduct is intentional or reckless, rather than merely negligent, is taken into account in determining whether it was the proximate cause of a given harm.[346] Proximate causation may be found in cases of intentional or reckless misconduct even though it might not have been found if the defendant's conduct were merely negligent.[347] In particular, while foreseeability usually sets the upper limit of liability for negligence,[348] a person may be held liable even for unintended and unforeseeable consequences of his intentional conduct.[349]

Imposing liability for unforeseeable harm seems superfluous from an efficient-deterrence perspective because the conduct of a potential injurer cannot be affected *ex ante* by unforeseeable liability for unforeseeable consequences.[350] The primary justification for imposing liability for unforeseeable consequences in cases of intentional wrongdoing thus seems to be retributive. Tort law enables courts to expand liability when they sense that applying ordinary principles of proximate causation may give rise to an abominable disproportion between the severity of the sanction and the gravity of the wrong. In this respect, relaxation of the proximate cause requirement operates much like punitive damages. Clearly, an award of punitive damages is

---

346. Under the assumption that the wrongful conduct was a cause-in-fact of the same harm.

347. United Food & Commercial Workers Unions, Employers Health & Welfare Fund v. Philip Morris, Inc., 223 F.3d 1271, 1274 (11th Cir. 2000); Meyers v. Epstein, 282 F. Supp. 2d 151, 154 (S.D.N.Y. 2003); Rodopoulos v. Sam Piki Enters., Inc., 570 So. 2d 661, 666 (Ala. 1990); Kowal v. Hofher, 436 A.2d 1, 3 (Conn. 1980); Kimberlin v. DeLong, 637 N.E.2d 121, 126-28 (Ind. 1994); PROSSER & KEETON, *supra* note 92, at 37; RESTATEMENT (SECOND) OF TORTS §§ 435B, 501(2) (1965). *Cf.* RESTATEMENT (THIRD) OF TORTS: LIABILITY FOR PHYSICAL HARM (Draft No. 3) § 33(b) (2003) (stating that a person "who intentionally or recklessly causes physical harm is subject to liability for a broader range of harms than the harms for which [he] would be liable if acting negligently.").

348. The wrongdoer is "not liable for harm different from the harms whose risks made the [wrongdoer's] conduct tortious." RESTATEMENT (THIRD) OF TORTS: LIABILITY FOR PHYSICAL HARM (Draft No. 3) § 29 (2003). In cases of negligence it is clear that the only risks whose creation may make one's conduct negligent are foreseeable risks. *Id.* § 29 cmt. d. *See also* RESTATEMENT (THIRD) OF TORTS: LIABILITY FOR PHYSICAL HARM (Draft No. 1) § 3 (2001) (stating that primary factors in ascertaining whether the creation of a certain risk is wrongful are the foreseeable probability that harm may ensue and the foreseeable severity of the harm that may ensue).

349. Caudle v. Betts, 512 So. 2d 389, 392 (La. 1987) (finding that where the wrong is intentional, the defendant may be held responsible even for unintended unforeseeable harms); *accord* Mayer v. Hampton, 497 A.2d 1206, 1209-10 (N.H. 1985); Seidel v. Greenberg, 260 A.2d 863, 872-74 (N.J. Super. Ct. Law Div. 1969); Baker v. Shymkiv, 451 N.E.2d 811, 813 (Ohio 1983); PROSSER & KEETON, *supra* note 92, at 40.

350. SHAVELL, *supra* note 10, at 129-30.

a better vehicle for the advancement of retributive justice, since the extent of punitive damages is far more flexible than the extent of liability based on actual loss. However, relaxation of the proximate cause requirement may be preferred by the judiciary because it appears to operate within the corrective structure of tort law.[351]

B. Protecting the Victim from an Undue Burden

Generally, tort law does not allow a completely innocent victim to bear her loss if it was caused by the wrongful conduct of another.[352] This principle derives from, and is therefore compatible with, the notion of corrective justice that underlies tort law. Yet, in most cases it also produces a just result from a retributive perspective (focusing on the victim): the wholly innocent victim is not punished by having to bear the loss.

A practical problem may arise where an innocent victim cannot recover from the tortfeasor because the latter is impecunious, legally immune, or untraceable. Tort law cannot resolve the ensuing retributive, and corrective, injustice where there is only one tortfeasor. However, it does seem to address the retributive concern in multiple-tortfeasor settings in which at least one tortfeasor can be sued and can pay damages. Under the well established principle of joint-and-several liability, if several persons act in concert to commit a tort, the victim may recover the full amount of the damages from any one or any combination of them.[353] More importantly, when independent negligent actions of several persons are each a proximate cause of an indivisible injury, the victim, may sue any one or any combination of them to obtain full recovery, at least when certain preconditions are met.[354] Clearly, the principle of joint-and-several liability may be justified in terms of corrective justice.[355] Yet another plausible and very common justification for this principle is that it prevents a completely innocent person from being burdened by an

---

351. Liability corresponds to actual losses caused, in the factual sense, by the defendant's conduct.

352. However, as shown above, retributive concerns may result in exclusion of liability where allowing recovery by innocent victims may give rise to an abominable disproportion between the gravity of the defendant's conduct and the severity of the aggregate sanction imposed on him. *See supra* Section IV.A.

353. Am. Motorcycle Ass'n v. Superior Court, 578 P.2d 899, 904 (Cal. 1978); RESTATEMENT (SECOND) OF TORTS §§ 875, 876 (1979); Paul Bargren, *Joint and Several Liability: Protection for Plaintiffs*, 1994 WIS. L. REV. 453, 455; Richard W. Wright, *The Logic and Fairness of Joint and Several Liability*, 23 MEM. ST. U. L. REV. 45, 70 (1992) [hereinafter Wright, *Joint and Several Liability*].

354. *Am. Motorcycle Ass'n*, 578 P.2d at 906-07; RESTATEMENT (SECOND) OF TORTS §§ 875, 876 (1979); Bargren, *supra* note 353, at 455-56; Wright, *Joint and Several Liability*, *supra* note 353, at 71.

355. Wright, *Joint and Several Liability*, *supra* note 353, at 54-62 (noting that each defendant whose tortious behavior was an actual and proximate cause of the injury is an actual and proximate cause of 100% of the injury, and is therefore individually fully responsible for the entire injury).

uncompensated harm.[356] Put differently, the principle of joint-and-several liability prevents an abominable disproportion between the (non-)culpability of the victim and the gravity of the burden she may incur in its absence. This is definitely a retributive formulation.[357]

One may argue that the principle of joint-and-several liability may give rise to an abominable disproportion between the gravity of the *defendant's* conduct and the severity of the sanction that he incurs.[358] This argument is unconvincing for two reasons. First, in many cases the doctrine only shifts to the defendant the burden of tracing the other injurers and litigating against them. It guarantees that the innocent plaintiff will not be "punished," but does not necessarily make the singled-out defendant bear the entire loss. Second, even if a single defendant in a multiple-tortfeasor setting is obliged to bear the entire loss, the discrepancy between the gravity of the wrong and the severity of the sanction is comparable to the typical disproportion in a single-tortfeasor setting. In the ordinary case, the defendant is "penalized" for wrongful conduct that was the actual and proximate cause of the plaintiff's harm or injury, and the severity of the penalty is determined by the extent of the actual loss. The principle of joint-and-several liability does not change this reality. A person whose wrongful conduct was the actual and proximate cause of the harm is obliged to pay for it.[359] True, disproportion may ensue. But ordinary (as opposed to abominable) disproportion cannot be taken into account within a predominantly corrective mechanism, in which the extent of liability is usually determined by the fortuitous magnitude of the actual loss.

VI. CONCLUSION

This article attempts to fill a gap in tort theory. Nowadays most theorists perceive tort doctrine as resting on either the notion of corrective justice, or on utilitarian-economic grounds. The concept of retributive justice remains outside the fierce controversy. This is quite understandable. Tort law—as a bipolar rectificatory mechanism—cannot attain retributive justice, nor can it be expected to. It frequently imposes sanctions on non-culpable parties; it does not impose sanctions on all wrongdoers; the extent of a sanction is usually determined by the magnitude of actual loss and cannot be adjusted to fit the

356. *Am. Motorcycle Ass'n*, 578 P.2d at 905 (noting that in some cases if joint and several liability were to be abandoned "a completely faultless plaintiff, rather than a wrongdoing defendant, would be forced to bear a portion of the loss if any one of the concurrent tortfeasors should prove financially unable to satisfy his proportioned share of the damages"); *accord* Coney v. J.L.G. Indus., Inc., 97 Ill. 2d 104, 122 (1983); Parks v. Union Carbide Corp., 602 S.W.2d 188, 200 (Mo. 1980).

357. The principle of joint-and-several liability is unique in that it seems to conform to both corrective and retributive considerations.

358. *See, e.g.*, McIntyre v. Balentine, 833 S.W.2d 52, 58 (Tenn. 1992) (noting that the principle of joint-and-several liability "may fortuitously impose a degree of liability that is out of all proportion to fault.").

359. The moral gravity of the wrong is not reduced by its co-existence with other wrongs.

gravity of the respective wrong, and in the end, the burden is not necessarily borne by the actual wrongdoer.

Nonetheless, the notion of retributive justice seems to have exerted a limited influence on the development of tort doctrine. The common law of torts is responsive to retributive concerns in two ways. First, it strives to prevent *abominable* disproportion between the severity of the sanction imposed on one of the litigating parties, and the gravity of his or her conduct. Second, it attempts to vindicate criminal justice as prescribed by criminal law where possible. These two paradigms have various manifestations.

Courts are inclined to limit or exclude liability where allowing full recovery would expose the defendant to an abominably excessive sanction in light of the gravity of his or her conduct. This inclination is apparent in the field of relational loss, or *dommage par ricochet*. Moreover, courts are likely to limit or exclude liability where allowing full recovery may enable the plaintiff to evade punishment for serious misconduct, or stultify his or her criminal liability. This has been done especially through the *ex turpi causa* defense. Courts also utilize retributive arguments to expand liability when ordinary principles of tort law impose an extremely lenient sanction on a culpable defendant. The foremost example, of course, is punitive damages. However, it has been argued that a similar rationale may explain the relaxation of the cause-in-fact and proximate cause requirements in certain types of cases. Finally, retributive concerns are brought into play where the defendant's liability is expanded to ensure that an innocent plaintiff will not have to bear any part of her loss.

To conclude, this article has had a descriptive purpose only. It has concentrated on the actual utilization of retributive concerns in tort jurisprudence and not on their precise normatively defensible role. The latter issue must await further research.

COPYRIGHT MISUSE v. FREEDOM OF CONTRACT: AND THE WINNER IS...

I. INTRODUCTION

The Copyright Act gives authors limited rights against the world. It grants authors the exclusive rights to reproduce, distribute, and prepare derivative works based on their copyrighted works.[1] The Act carefully strikes a balance between incentivizing authors to create works and assuring public access to those works.[2] This balance is achieved, in part, by giving authors special rights of Copyright for "limited [t]imes" before their entire work becomes part of the public domain.[3] The balance is furthered by limiting the copyright grant's coverage; for example, an author's mode of expression is protected while the ideas underlying her work immediately enter the public domain.[4] Finally, the balance is maintained by explicit statutory exceptions to the author's exclusive rights,[5] such as fair use and the first sale doctrine.[6]

Authors increasingly use contracts to circumvent the statutory exceptions to their exclusive rights under the Copyright Act.[7] Authors use such contracts, often referred to as licenses, to structure agreements that give them more rights with respect to the licensee than the general Copyright Act gives them with respect to the world. The licensee agrees to forego these statutory exceptions presumably because the protections they provide have less value to the individual licensee than access to the copyrighted work.[8]

What contract provisions are appropriate (and enforceable) between an author/licensor and a licensee? The answer depends on whether one views the question from (1) a copyright misuse perspective or (2) a freedom of contract

---

1. The Copyright Act of 1976, 17 U.S.C. § 106 (2000).

2. *See, e.g.*, Brett Frischmann & Dan Moylan, *The Evolving Common Law Doctrine of Copyright Misuse: A Unified Theory and Its Application to Software*, 15 BERKELEY TECH. L.J. 865, 901 (2000) [hereinafter Frischmann & Moylan] ("The public policy behind the copyright system *is premised* upon an exchange between short-term 'monopoly' costs and long-term efficiency gains in investment, production, and dissemination of innovation.").

3. U.S. CONST. art. 1, § 8, cl. 8.

4. 17 U.S.C. § 102(b) (2000) ("In no case does copyright protection for an original work of authorship extend to any idea . . . .").

5. *See* 17 U.S.C. §§ 107-122 (2000).

6. *See* 17 U.S.C. §§ 107, 109 (2000).

7. *See, e.g.*, Lydia Pallas Loren, *Slaying the Leather-Winged Demons in the Night: Reforming Copyright Owner Contracting with Clickwrap Misuse*, 30 OHIO N.U. L. REV. 495, 495 (2004) ("In the era of digital delivery of content, copyright owners have turned with a vengeance to the institution of contract to specify the rights and responsibilities of their customers.").

8. *See, e.g.*, Kathryn Judge, Note, *Rethinking Copyright Misuse*, 57 STAN. L. REV. 901, 938-39 (2004) (discussing the relatively low value of the right to reverse engineer software or criticize a book to an individual as compared to the value of such rights to the society at large).

perspective. The copyright misuse perspective distrusts attempts to expand the constitutionally limited monopoly granted to authors under copyright law.[9] The freedom of contract perspective supports an individual's right to contract as he sees fit.[10] Five federal circuit courts have explicitly adopted the Copyright Misuse Doctrine, while finding copyright misuse in three cases involving a copyright holder's contractually expanded rights.[11] Four circuits have upheld the individual's right to contract around the statutory exceptions in the Act.[12] These courts either did not discuss copyright misuse[13] or did not find copyright misuse applicable to a breach of contract claim.[14] Of those circuits, only one has adopted the copyright misuse doctrine.[15]

This Comment compares cases decided under a copyright misuse rubric with those decided under a freedom of contract rubric. It examines the outcomes of those cases and the courts' varying rationales. Part II explores the current status of the copyright misuse doctrine and includes a brief review of the key circuit court cases adopting the copyright misuse doctrine. Part II also analyzes the primary justifications for and proposed developments of the copyright misuse doctrine. Part III explores cases driven by a freedom of contract perspective as well as commentaries on these cases. Part IV compares a case decided from the copyright misuse perspective with a case decided from a freedom of contract perspective in an effort to explore the courts' differing analyses. Finally, Part IV(c) discusses combining some of the best points from

---

9. *See, e.g.*, Lasercomb Am., Inc. v. Reynolds, 911 F.2d 970, 977 (4th Cir. 1990) (stating that copyright misuse "equally forbids the use of the [copyright] to secure an exclusive right or limited monopoly not granted by the [Copyright] Office and which it is contrary to public policy to grant" (quoting Morton Salt Co. v. G.S. Suppiger, 314 U.S. 488, 492 (1942) (brackets in original))).

10. Bowers v. Baystate Techs., Inc., 320 F.3d 1317, 1323 (Fed. Cir. 2003) (stating that "[c]ourts respect freedom of contract and do not lightly set aside freely-entered agreements").

11. Assessment Techs. of Wis., LLC v. WIREdata, Inc., 350 F.3d 640, 647 (7th Cir. 2003) (adopting the copyright misuse doctrine but finding no copyright misuse); Video Pipeline, Inc. v. Buena Vista Home Entm't, Inc., 342 F.3d 191, 203-06 (3d Cir. 2003) (finding no copyright misuse); Practice Mgmt. Info. Corp. v. Am. Med. Ass'n, 121 F.3d 516, 520-21 (9th Cir. 1997) (finding copyright misuse); DSC Commc'ns Corp. v. DGI Techs., Inc. (DSC I), 81 F.3d 597, 601-02 (5th Cir. 1996) (while adopting the doctrine of copyright misuse in 1996, the circuit court overruled the judge's discretion and reinstated the jury finding of copyright misuse in *Alcatel USA, Inc. v. DGI Technologies., Inc.* (DSC II), 166 F.3d 772, 793-94 (5th Cir. 1999)); Lasercomb Am., Inc. v. Reynolds, 911 F.2d 970, 976-79 (4th Cir. 1990) (finding copyright misuse).

12. Altera Corp. v. Clear Logic, Inc., 424 F.3d 1079, 1090 (9th Cir. 2005); Davidson & Assocs. v. Jung, 422 F.3d. 630, 639 (8th Cir. 2005); Bowers v. Baystate Techs., Inc., 320 F.3d 1317 (Fed. Cir. 2003); ProCD, Inc. v. Zeidenberg, 86 F.3d 1447, 1450-55 (7th Cir. 1996).

13. *Bowers*, 320 F.3d at 1317; *ProCD*, 86 F.3d at 1447.

14. *Altera*, 424 F.3d at 1090; *Davidson*, 422 F.3d at 639.

15. The Ninth Circuit, having adopted the copyright misuse doctrine, has come to differing conclusions in two copyright misuse cases. *See Altera*, 424 F.3d at 1090; *Practice Mgmt.*, 121 F.2d at 520-21.

both perspectives into an analysis that properly balances the public's right to free information with the individual's right to contract.

A. Copyright Misuse

1. Court Cases

Copyright misuse is an equitable doctrine that denies copyright infringement relief to copyright holders who have misused their limited monopoly under the Copyright Act.[16] Copyright misuse, like patent misuse, does not render a copyright invalid; a finding of copyright misuse simply makes the copyright unenforceable while the offending behavior continues.[17] *Lasercomb America, Inc. v. Reynolds*[18] is often cited as the landmark case for copyright misuse.[19] The key circuit cases following *Lasercomb* have analyzed license agreements for copyright misuse.[20] Judge Posner, in his opinion on *Assessment Technologies of Wisconsin, LLC v. WIREdata, Inc.*,[21] discussed an abuse of process perspective on copyright misuse: an author using the threat of a copyright suit to discourage a defendant from exercising his legitimate rights under the Copyright Act (e.g., fair use) may be subject to a ruling of copyright misuse.[22]

a. Contract-based Copyright Misuse Cases - *Lasercomb / Alcatel (DSC II) / Practice Management / Video Pipeline*

The Fourth Circuit's decision in *Lasercomb America, Inc. v. Reynolds*[23] is often cited as the case that brought the copyright misuse defense into its modern form.[24] Lasercomb licensed its computer software under a ninety-nine-

16. *See, e.g., Video Pipeline*, 342 F.3d at 204 ("The misuse doctrine extends from the equitable principle that courts 'may appropriately withhold their aid where the plaintiff is using the right asserted contrary to the public interest.'" (quoting Morton Salt Co. v. G.S. Suppiger Co., 314 U.S. 488, 492 (1942))).

17. *See, e.g., id.* ("Misuse is not cause to invalidate the copyright or patent but instead 'precludes its enforcement during the period of misuse.'" (quoting *Practice Mgmt. Info. Corp.*, 121 F.3d at 520 n.9)); *Lasercomb*, 911 F.2d at 979 n.22 ("This holding . . . is not an invalidation of Lasercomb's copyright. Lasercomb is free to bring a suit for infringement once it has purged itself of the misuse.").

18. 911 F.2d 970 (4th Cir. 1990).

19. *See, e.g., Video Pipeline, Inc.*, 342 F.3d at 203-06; Frischmann & Moylan, *supra* note 2, at 888 ("[T]he [*Lasercomb*] decision was the first affirmative application of the copyright misuse defense in a federal court of appeals.").

20. *See Video Pipeline*, 342 F.3d at 203-06; Alcatel USA, Inc. v. DGI Techs., Inc. (DSC II), 166 F.3d 772, 792-95 (5th Cir. 1999); *Practice Mgmt. Info. Corp.*, 121 F.3d at 520-21; DSC Commc'ns Corp. v. DGI Techs., Inc. (DSC I), 81 F.3d 597, 601 (5th Cir. 1996).

21. 350 F.3d 640 (7th Cir. 2003).

22. *Id.* at 647.

23. 911 F.2d 970 (4th Cir. 1990).

24. *See, e.g.*, Frischmann & Moylan, *supra* note 2, at 888 ("[T]he [*Lasercomb*] decision

year agreement that forbade licensees "to write, develop, produce or sell computer assisted die making software."[25] The Fourth Circuit's opinion considered the historical development of patent and copyright law and then adopted a copyright misuse defense parallel to the patent misuse defense articulated by the U.S. Supreme Court in *Morton Salt Co. v. G.S. Suppiger Co.*[26]

> The grant to the [author] of the special privilege of a [copyright] carries out a public policy adopted by the Constitution and laws of the United States, 'to promote the Progress of Science and the useful Arts, by securing for limited Times to [Authors] . . . the exclusive Right . . .' to their ["original" works]. United States Constitution, Art. I, § 8, cl. 8, [17 U.S.C.A. § 102]. But the public policy which includes [original works] within the granted monopoly excludes from it all that is not embraced in the [original expression]. It equally forbids the use of the [Act] to secure an exclusive right or limited monopoly not granted by the [Copyright] Office and which it is contrary to public policy to grant.[27]

The Fourth Circuit rejected the idea that the copyright misuse defense could only apply if Lasercomb's actions violated antitrust laws and directly affected the defendant.[28] "The misuse arises from Lasercomb's attempt [through its standard license agreement] to use its copyright . . . to control competition in an area outside the copyright, *i.e.*, the idea of computer-assisted die manufacture, regardless of whether such conduct amounts to an antitrust violation."[29]

---

was the first affirmative application of the copyright misuse defense in a federal court of appeals."); Judge, *supra* note 8, at 916 ("The first federal appellate court to apply the doctrine of copyright misuse was the Fourth Circuit, in the case of *Lasercomb America, Inc. v. Reynolds*.").

25. *Lasercomb*, 911 F.2d at 973.

26. 314 U.S. 488 (1942). Criticizing the Fourth Circuit's reliance on the *Morton Salt* case, Jeffrey Andrews wrote the following:

> It is interesting to note that although the Fourth Circuit relied heavily on the similarities between the justifications for patent and copyright protection, it ignored the recent developments shifting the scope of patent misuse toward antitrust and away from public policy. . . . [T]he court in *Lasercomb* ignored considerations of market power and based its holding on the plain language of the licensing agreement rather than on the actual market effect.

Jeffery A. Andrews, Comment, *Reversing Copyright Misuse: Enforcing Contractual Prohibitions on Software Reverse Engineering*, 41 HOUS. L. REV. 975, 997 (2004).

27. *Lasercomb*, 911 F.2d at 977 (quoting with alterations *Morton Salt*, 314 U.S. at 492).

28. *Id.* at 978-79.

29. *Id.* at 979. *See also* Neal Hartzog, *Gaining Momentum: A Review of Recent Developments Surrounding the Expansion of the Copyright Misuse Doctrine and Analysis of the Doctrine in Its Current Form*, 10 MICH. TELECOMM. & TECH. L. REV. 373, 384-85 (2004) ("Therefore, the court in *Lasercomb* explicitly adopted a 'scope of the grant' copyright misuse analysis (i.e. the 'public policy' analysis) . . . . It is important to note that the court focused on the unreasonableness of the licensing agreement, and essentially ignored 'the actual effects on competition or market power of the plaintiff, as it would in an antitrust analysis.'" (quoting

The Ninth, Fifth, and most recently, Third Circuit Courts have explicitly adopted the copyright misuse defense as defined by the Fourth Circuit in *Lasercomb*.[30] In *Practice Management Information Corp. v. the American Medical Ass'n*,[31] Practice Management sought a declaratory judgment that the American Medical Association's (AMA) Physician's Current Procedural Terminology (CPT) had become uncopyrightable, or alternatively, that the AMA had misused its copyright.[32] The Ninth Circuit, with little additional reasoning, adopted *Lasercomb* and declared that the AMA had misused its copyright by "[c]onditioning the license on [the government agency's] promise not to use competitors' products."[33]

The Fifth Circuit adopted the copyright misuse doctrine when it upheld a preliminary injunction in *DSC Communications Corp. v. DGI Technologies (DSC I)*.[34] DSC brought a copyright infringement suit against DGI, a competitor that used DSC's software to develop competitive phone switches (microprocessors).[35] DSC licensed its operating software for use "in conjunction with the phone switch[es] purchased from DSC."[36] The Fifth Circuit explicitly adopted the copyright misuse doctrine from *Lasercomb*, stating, "[w]e concur with the Fourth Circuit's characterization of the copyright misuse defense."[37] The Fifth Circuit noted that

> DSC seems to be attempting to use its copyright to obtain a patent-like monopoly over unpatented microprocessor cards. . . . The defense of copyright misuse "forbids the use of the copyright to secure an exclusive right or limited monopoly not granted by the Copyright Office," including a limited monopoly over microprocessor cards.[38]

Frischman & Moylan, *supra* note 2, at 890)).

Kathryn Judge analyzed three ways "of understanding when a copyright holder is engaged in misuse." *See* Judge, *supra* note 8, at 922-23. (1) "copyright misuse arises whenever a copyright holder attempts to use his copyright in an anticompetitive fashion . . ." (2) "misuse arises whenever a copyright holder attempts to extend the scope of his exclusive rights beyond the formal bounds created by the legislative scheme," and (3) "misuse arises when a copyright holder crosses certain lines that are particularly central to copyright policy." *Id.* at 917.

30. *See* cases cited *supra* note 11.

31. 121 F.3d 516 (9th Cir. 1997).

32. *Id.* at 518.

33. *Id.* at 521 (citing *Lasercomb*, 911 F.2d at 977).

34. 81 F.3d 597, 600-01 (5th Cir. 1996).

35. *Id.* at 598-99 (5th Cir. 1996).

36. *Id.* at 599.

37. *Id.* at 601.

38. *Id.* at 601 (quoting *Lasercomb*, 911 F.2d at 977). *See also* Judge, *supra* note 8, at 920 ("Fair use by its nature is a very fact-intensive inquiry, and its applicability is determined by reference to the actions of the alleged infringer. Accordingly, a finding of fair use may have protected DGI, but it would have provided minimal comfort to others seeking to compete with Alcatel and it would have had no effect on discouraging Alcatel from engaging in aggressive licensing practices.").

During the course of litigation, Alcatel USA, Inc. purchased DSC.[39] After trial, when the case reached the Fifth Circuit for the second time as *Alcatel USA, Inc. v. DGI Technologies, Inc. (DSC II)*,[40] the Fifth Circuit held that the district court had abused its discretion in overturning the jury finding of copyright misuse.[41] The circuit court overruled any parts of the injunction related to DGI's infringement; the copyright was unenforceable because DSC had misused its copyright.[42]

Through *Video Pipeline, Inc. v. Buena Vista Home Entertainment, Inc.*,[43] the Third Circuit became the most recent circuit to adopt the copyright misuse doctrine as defined in *Lasercomb*.[44] Video Pipelines did not have a license to show Disney trailers on its website; instead, Video Pipeline produced its own trailers from Disney films.[45] Video Pipelines sought a declaratory judgment that it was not violating Disney's copyrights, and Disney countersued for copyright infringement.[46]

Disney licensed use of its movie trailers to websites under explicit restrictions: "The Website[s] . . . may not be derogatory to or critical of the entertainment industry or of [Disney] . . . or of any motion picture produced or distributed by [Disney] . . . ."[47] Video Pipeline asserted that Disney's "licensing agreements seek to use copyright law to suppress criticism and, in so doing, . . . trigger[] the copyright misuse doctrine."[48] The Third Circuit briefly reviewed the patent misuse and copyright misuse doctrines[49] and "extend[ed] the patent misuse doctrine to copyright, and recognize[d] that it might operate beyond its traditional anti-competition context . . . ."[50] The Third Circuit, however, did not consider the licensing agreements to "interfere with creative expression to such a degree that they affect in any significant way the policy interest in increasing the public store of creative activity."[51] Additionally, the court expressed concern that Disney might display its trailers only on its own

---

39. Alcatel USA, Inc. v. DGI Techs. Inc. (DSC II), 166 F.3d 772, 777 (5th Cir. 1999).

40. 166 F.3d 772 (5th Cir. 1999).

41. *Id.* at 793-95.

42. *Id.* at 777. The circuit court upheld parts of the injunction and damages based on DGI's misappropriation of trade secrets. *Id.* at 799. The court "vacate[d] all legal and equitable relief awarded to DSC" on its unfair competition claim because it was preempted by the Copyright Act. *Id.* As the injunction was based upon "DGI's trade secret misappropriation, unfair competition by misappropriation, . . . and copyright infringement," the case was remanded so that the district court could "revise its injunction if and to the extent the court deems necessary or desirable." *Id.* at 789, 799.

43. 342 F.3d 191 (3d Cir. 2003).

44. *Id.* at 203-06.

45. *Id.* at 195.

46. *Id.* at 196.

47. *Id.* at 203 (quoting language from Disney's trailer licensing agreements) (brackets in original).

48. *Id.*

49. *Id.* at 203-06

50. *Id.* at 206.

51. *Id.*

website if its alternative was to license to other websites "willy-nilly regardless of the content displayed with its copyrighted works."[52] In other words, copyright policy's important goal of maximizing public access to creative works was best served by a finding that Disney's license agreement did not constitute copyright misuse.[53]

Taken together, these cases can be distilled into this frequently quoted statement of proof for a copyright misuse defense: "To establish copyright misuse, a defendant must establish either '(1) that [the plaintiff] violated the antirust laws, or (2) that [the plaintiff] illegally extended its monopoly beyond the scope of the copyright or violated the public policies underlying the copyright laws.'"[54] In other words, a defendant may establish copyright misuse by showing (1) that the plaintiff violated the antitrust laws, (2) that the plaintiff illegally extended its monopoly beyond that of the copyright, or (3) that the plaintiff otherwise violated the public policies underlying the copyright laws. Arguably, the authors' misuses in *Lasercomb*, *Practice Management*, and *DSC II* were based upon the second prong because in each of those cases the copyright holder attempted to extend its copyright beyond the scope of the Act.[55] The issue in *Video Pipeline* can be articulated under the third prong: whether the license agreement violates the fair use public policy of copyright law.[56]

b. Abuse-of-process-based Copyright Misuse—*WIREdata*

In the previous cases, courts applied the copyright misuse doctrine solely to license agreements. Judge Posner, in spite of originally arguing that the scope of copyright misuse should be limited to antitrust violations,[57] now suggests that procedural issues may lead to a finding of copyright misuse.[58] In

52. *Id.*

53. *Id.* ("Indeed [a contrary] application of the misuse doctrine would likely decrease the public's access to Disney's works because it might as a result refuse to license at all online displays of its works."); *see also* Loren, *supra* note 7, at 533 ("[T]he court weighed the widespread availability of Disney products, restriction free, together with the limited context in which the restrictive clauses were used to conclude that those clauses did not constitute misuse.").

54. Lexmark Int'l, Inc. v. Static Control Components, Inc., 253 F. Supp. 2d 943, 966 (E.D. Ky. 2003) (quoting Microsoft Corp. v. Compusource Distribs., Inc., 115 F. Supp. 2d 800, 811 (E.D. Mich. 2000) (brackets in original)).

55. *See, e.g.*, Frischmann & Moylan, *supra* note 2, at 901 (stating that these cases relied on the actions of the copyright holders rather than the effect of those actions on the public).

56. *See* Loren, *supra* note 7, at 519 ("[T]he Third Circuit's acceptance of a broad copyright misuse doctrine, informed by public policy beyond antitrust concerns, should give copyright owners pause . . . to assure that their licens[ing] agreements do not restrict critical expression . . . .").

57. *See* Saturday Evening Post Co. v. Rumbleseat Press, Inc., 816 F.2d 1191, 1200 (7th Cir. 1987).

58. *See* Assessment Techs. of Wis., LLC v. WIREdata, Inc., 350 F.3d 640, 645-47 (7th Cir. 2003); William F. Patry & Richard A. Posner, *Fair Use and Statutory Reform in the Wake of* Eldred, 92 CAL. L. REV. 1639,1654-59 (2004).

Assessment Technologies of Wisconsin, LLC v. WIREdata, Inc.,[59] Judge Posner, apparently *sua sponte*, discussed Assessment Technologies' (AT) potential copyright misuse.[60] In that case, AT held a valid copyright in a software program, Market Drive, that allowed "[t]he municipality's tax officials [to] use various queries . . . to view the data in the file."[61] City employees, not AT, collected the information for property-tax purposes and inputted the raw data into the file.[62] When the Southeastern Wisconsin municipalities refused to furnish the data to WIREdata in fear of violating AT's copyright, WIREdata sued the municipalities.[63] In response, AT sued WIREdata for copyright infringement.[64] Judge Posner noted that

> The argument for applying copyright misuse . . . is that for a copyright owner to use an infringement suit to obtain property protection, here in data, that copyright law clearly does not confer, hoping to force a settlement or even achieve an outright victory over an opponent that may lack the resources or the legal sophistication to resist effectively, is an abuse of process.[65]

This aspect of copyright misuse is an interesting equitable doctrine that could give the public greater freedom with regard to its legitimate use of copyrighted and uncopyrightable materials.[66] This Comment, however, focuses on contract-based copyright misuse and bases its comparison of court analysis and outcome on that doctrine.

2. The Case for Copyright Misuse

Several commentators have developed theories of copyright misuse that

---

59. 350 F.3d 640 (7th Cir. 2003).

60. *Id.* at 645-47.

61. *Id.* at 643.

62. *Id.*

63. *Id.* at 642. WIREdata wanted the property data, which included addresses, owner information, and valuations for use by real estate brokers. *Id.*

64. *Id.*

65. *Id.* at 647. Judge Posner opened his opinion with the following statements:

> This case is about the attempt of a copyright owner to use copyright law to block access to data that not only are neither copyrightable nor copyrighted, but were not created or obtained by the copyright owner. The owner is trying to secrete the data in its copyrighted program–a program the existence of which reduced the likelihood that the data would be retained in a form in which they would have been readily accessible. It would be appalling if such an attempt could succeed.

Id. at 641; *see also*, Judge, *supra* note 8, at 928-29 (stating that Patry and Posner see the copyright misuse doctrine as providing a safe harbor for "'situations in which asymmetrical stakes discourage a legal challenge to the claim'") (quoting Patry & Posner, *supra* note 58, at 1658).

66. *See* Judge, *supra* note 8, at 933. (stating that if the procedural copyright misuse doctrine were applied with uniform objective rules, the public would know that copyright misuse had occurred without having to seek a declaratory judgment to that effect).

describe how the courts have developed and should continue to develop the doctrine.[67] These commentators tend to base the need for a copyright misuse doctrine on the inability of copyright to effectively regulate software[68] or on the aggressive licensing schemes utilized by today's copyright owners.[69] Following the discussion on the need for a copyright misuse doctrine are suggestions by commentators on how to better define and apply the doctrine.

a. Why Copyright Misuse is Needed

The arguments in favor of the copyright misuse doctrine revolve around the following three points: copyright protection is constitutionally limited,[70] copyright limitations are more valuable to society as a whole than to individuals,[71] and copyright misuse focuses the court's attention on the copyright holder's behavior.[72] Copyright and patent protections exist in tension with our open, competitive market beliefs. They offer a limited monopoly in a legal framework that otherwise abhors monopolies. The Constitution sets express limits on the monopoly as to time and material covered.[73] As a society, we want to limit the scope and length of a monopoly to the absolute minimum that is necessary to incentivize authors to create. As such, "[t]he Copyright Act . . . expressly limits the rights granted to copyright owners in 16 separate sections of the statue."[74] When authors attempt to expand their rights beyond this government granted monopoly, the carefully legislated balance between an author's rights to her work and the public's right to information is disrupted.

Many commentators argue that, aside from contracting, the rights of copyright owners have also increased in recent years because "Congress is

67. *See, e.g.,* Frischmann & Moylan, *supra* note 2; Loren, *supra* note 7; Judge, *supra* note 8.

68. *See* Frischmann & Moylan, *supra* note 2, at 912 ("Computer software is currently protected through copyright because both source code and object code are considered 'written expressions.' Yet the economic value of copyrighted object code is completely derived from the functional ends facilitated by the software. . . . Correspondingly, copyright misuse is an appropriate judicial mechanism for restricting the social costs of granting copyrights on functional innovations."). *But see* Judge, *supra* note 8, at 914-15 (rejecting the copyright misuse /software connection because she questions its accuracy and "because of doubts about whether the functional nature of software and the complications that this creates for copyright policy are issues that can be rectified by copyright misuse").

69. *See, e.g.,* Judge, *supra* note 8, at 915 (stating that she is "inclined to believe that it was actually aggressive licensing practices that instigated the doctrine's adoption; and, more importantly, that it is such abuses the doctrine should seek to curb"); Loren, *supra* note 7, at 499-500 (arguing that "the current legal doctrines available to invalidate these overreaching [contract] provisions or to strike claims asserted for their breach are insufficient to stem the tide of abusive contracting behavior by content owners").

70. *See* Loren, *supra* note 7, at 498.

71. *See id.* at 534.

72. *See* Judge, *supra* note 8, at 915.

73. U.S. CONST. art. I, § 8, cl. 8.

74. Loren, *supra* note 7, at 498.

armed with only a one-way ratchet" with regard to copyright legislation.[75] They further argue that the extension of copyright protection to software, which is both expressive and functional, and the growing use of standardized contracts, have significantly impacted copyright holders' inherent power.[76] The copyright misuse doctrine is especially important because it helps to constrain copyright owners from further expanding their rights beyond those granted in the Act.[77]

The copyright holder has limited statutory rights against the world. As previously discussed, these limited rights strike a careful balance between limited monopoly and the incentive to create. Society as a whole benefits from this balance. The copyright holder and a specific individual may value a different balance of rights with respect to one another than the balance of statutorily created rights between the copyright holder and the world. Advocates of the copyright misuse doctrine worry that an individual's costs in contractually waiving a given right are low compared to the costs for society.[78] This is because many of the uses involved in restricting the copyright holder's rights against the world are uses "that have external benefits that are unlikely to be internalized by the party that has contracted away the right to engage in such uses."[79] For example, criticism, a traditional fair use of a copyrighted work, has a much greater value to society as a whole than to a given individual; society benefits more from access to the criticisms than any one individual does from criticizing.[80] Likewise, only a few users of computer software would benefit from using fair-use reverse engineering to reveal the software's uncopyrightable ideas and develop competing products; society, however, benefits greatly from this competition.[81] Contracts in which individuals give up rights in favor of the copyright holder, when broadly or repeatedly applied, not only impact the contracting parties but also effectively serve as private legislation changing the balance of power.[82]

---

75. Judge, *supra* note 8, at 913; *see also id.* at 906-07 (enumerating various expansions of copyright holder rights from the 1976 Copyright Act through the Digital Millennium Copyright Act).

76. *See, e.g.*, Frischmann & Moylan, *supra* note 2, at 907 ("Increasingly, producers supplement their copyright protection of hidden expression through alternate frameworks like licensing restrictions or technology in order to raise the costs of reverse engineering."); Loren, *supra* note 7, at 530 ("Indeed, at the time the current Copyright Act was passed, no one envisioned a distribution technology that would permit copyright owners to effectively enter into contracts with potentially *all* users of their works."); Judge, *supra* note 8, at 907.

77. Loren, *supra* note 7, at 519 ("As the rights of copyright owners become more powerful, it is appropriate for courts, using their equitable powers, to place more burdens on the exercise of those rights to minimize the risk of abuse of those powerful rights.").

78. *See id.* at 530.

79. *Id.* at 534.

80. *See, e.g., id.* at 504-505.

81. *See, e.g., id.* at 505-06 ("If the only way for a potential competitor to obtain a copy of the work includes binding themselves to a shrinkwrap license . . . that restriction [against reverse engineering] will enter into the competitor's determination of whether to proceed in making the competing product . . . .").

82. *See, e.g., id.* at 525 (arguing that misuse doctrine can be used "to reaffirm the limited

Finally, copyright misuse has been characterized as an excellent augmentation to the fair use doctrine.[83] Copyright misuse focuses on the behavior of the copyright holder while fair use focuses on the behavior of the alleged infringer.[84] The use of "overreaching contractual terms . . . in ubiquitous shrinkwrap and clickwrap licenses has a certain *in terrorem* effect on users."[85] The potential chilling effect on legitimate uses of copyrighted works is disconcerting. The courts are not using the doctrine of copyright preemption to address these issues; in fact, they are increasingly avoiding preemption findings.[86]

b. How Copyright Misuse Can Work Most Effectively

Advocates for the copyright misuse doctrine argue that it is a "tool for shaping the contracting behavior of intellectual property owners."[87] Brett Frischmann and Dan Moylan note in *The Evolving Common Law Doctrine of Copyright Misue: A Unified Theory and Its Application to Software* that neither the *Lasercomb*, *Practice Management*, nor *Alcatel (DSC II)* courts "have enunciated a form of public policy misuse that does not involve some degree of anticompetitive effects."[88] The authors conclude that "restricting a licensee's ability to innovate or tying an uncopyrighted or unpatented good to a copyrighted good is likely to be per se misuse."[89] The authors advocate creating a "small set of . . . per se rules," as in patent misuse; if a per se rule does not

---

scope of rights granted by Congress in the Copyright Act and to prevent copyright owners from attempting to re-legislate their rights customer by customer").

83. *See, e.g.*, Hartzog, *supra* note 29, at 404; *cf.* Loren, *supra* note 7, at 523 ("Commentators have recognized the interplay between fair use and misuse, particularly in the context of computer programs.").

84. *See, e.g.*, Judge, *supra* note 8, at 915 ("Copyright misuse is one of the only copyright-limiting doctrines that arises from actions taken by the copyright holder.").

85. Loren, *supra* note 7, at 500. "Shrinkwrap" licenses are "written [software] licenses that become effective as soon as the customer tears the wrapping from the package." ProCD, Inc. v. Zeidenberg, 86 F.3d 1447, 1449 (7th Cir. 1996).

86. *See, e.g.*, Bowers v. Baystate Techs., Inc., 320 F.3d 1317, 1325 (Fed. Cir. 2003); Loren, *supra* note 7, at 506-07 ("Despite widespread scholarly criticism of the decision, many courts have followed the reasoning of *ProCD* [finding an enforceable contract] in the clickwrap context.") (internal footnote omitted); *see also id.* at 510 ("When the Copyright Act provides the public policy at issue, courts typically will analyze the assertions that the clause is void because it violates public policy under the doctrine of federal preemption instead.").

87. *See, e.g.*, Loren *supra* note 7, at 519.

88. Frischmann & Moylan, *supra* note 2, at 901 ("Most courts evaluating copyright misuse defense acknowledge the equitable nature of the doctrine and the scope limitation function it plays, yet none have enunciated a form of public policy misuse that does not involve some degree of anticompetitive effects.").

89. *Id.* at 902. *See also* Andrews, *supra* note 26, at 994-1003 (reviewing *Lasercomb*, *Practice Management*, and *DSC II* and noting that each "opinion contained no discussion of market power"). Frischmann and Moylan suggest that a few narrow factual situations should constitute per se copyright misuse. Frischmann & Moylan, *supra* note 2, at 902.

apply, courts should consider copyright misuse "under a rule of reason analysis."[90] Computer software is especially well suited to a per se rule regarding reverse engineering for the following reasons: (1) the expressive elements of a program are not disclosed in the object code; (2) the ideas imbedded in the computer program are not disclosed; and (3) the software program's key source of value is in its functional elements, which are not copyrightable.[91] Copyright misuse protects the public interest in software by filling the non-disclosure gap in the copyright law and coordinating patent and copyright law.[92]

Lydia Pallas Loren proposes that courts use a rebutable presumption of copyright misuse if the terms of a contract expand the copyright owner's rights beyond the limitations of the Copyright Act.[93] The presumption would arise if the copyright owner explicitly contracts around the limits set in Copyright Act sections 107-122, or if the copyright owner attempts to protect noncopyrightable aspects of his work through contract.[94] Loren argues that a presumption of copyright misuse would deter copyright owners from overreaching in their

---

90. Frischmann & Moylan, *supra* note 2, at 902 ("Generally, copyright misuse should closely resemble patent misuse, having a small set of 'known' per se rules and otherwise being determined under a rule of reason analysis."). *See also* Hartzog, *supra* note 29, at 396 (suggesting that employment of a "'rule of reason' analysis" would "violate precedent" because "courts recognizing the misuse doctrine have done so utilizing a per se . . . analysis[.]").

91. *See* Frischmann & Moylan, *supra* note 2, at 905-914. Copyright misuse doctrine attempts to coordinate patent and copyright laws for computer software because software is copyrighted but has most of its value in its functional aspects. *Id.* at 876-77. *See also* Andrews, *supra* note 26, at 984-85 (stating that software manufacturers use a combination of object code distribution and copyright to gain protection while avoiding the disclosure necessary to obtain patent protection).

92. Frischmann & Moylan, *supra* note 2, at 878-80. Frischmann and Moylan argue that "[t]he hidden nature of software exploits a gap in the Copyright Act, namely the absence of an explicit requirement that expressive works be perceptible." *Id.* at 919. If the common law fails to fill this gap, "software producers can effectively acquire a monopoly of underlying ideas and expression, all the while maintaining secrecy in their designs, for a term of life plus seventy years." *Id.* at 921. Frischmann and Moylan suggest that courts can use a copyright misuse doctrine to "restore the public-private balance in copyright law." *Id.* at 927. Under a unified approach,

> (1) courts should apply traditional antitrust principles, consisting of per se rules and the rule of reason, when coordinating copyright and antitrust or gap-filling along the interface of these bodies of law; and (2) courts should avoid any form of reasonableness or balancing test and should constrain themselves to formulating per se rules when filling gaps in the copyright laws, coordinating copyright and patent law, or safeguarding the policies underlying the copyright laws.

Id. at 879. *See also* Andrews, *supra* note 26, at 980. ("[T]he issue of how to properly protect intellectual property rights embodied in software provides a challenge because, under current law, a single piece of software cannot be fully protected by either copyright or patent alone."); Hartzog, *supra* note 29, at 377 ("[T]he extension of copyright protection to software and technology has greatly increased the monopolistic opportunities inherent in the granting of a copyright.").

93. *See* Loren, *supra* note 7, at 522.

94. *See id.* at 524.

clickwrap license agreements that apply to each purchaser of the copyrighted material.[95] Such deterrence would ameliorate the corresponding negative social impact of this "private legislation" on the purposes behind copyright.[96] The copyright misuse doctrine would allow the defendant to use the defense and escape infringement liability even if the defendant was not negatively affected by the misuse.[97] Noting the courts' reluctance to allow an infringer to completely avoid liability, Loren argues that "courts should acknowledge the role of permitting infringers to escape liability in driving home the importance for copyright owners to remain within the bounds of the rights granted by the Copyright Act in their contracting behavior."[98] The copyright in a software program with a clickwrap license forbidding reverse engineering would likely be unenforceable under a copyright misuse defense.[99] Loren argues that such an outcome would discourage copyright owners from overreaching in their license agreements.[100]

In her Note entitled *Rethinking Copyright Misuse*, Kathryn Judge advocates using the copyright misuse doctrine to limit an offending copyright holder to relief for infringement damages instead of the equitable relief of copyright.[101] She notes that the "policy goal [of] a competitive market place, and [its] correspondent guideline, federal antitrust law, already has an established position in copyright misuse."[102] Judge argues that, in addition to a competitive market place, "copyright misuse should protect the values embodied in the First Amendment, and that it can do this if any attempt by a copyright holder to cross the idea/expression boundary or to deter fair use of his copyright material is deemed misuse."[103]

The copyright misuse cases and commentaries are fundamentally concerned with maintaining an appropriate balance of rights between the copyright holder and the public. Attempts by a software licensor to block all reverse engineering uses of a computer program, for example, would be inconsistent with the goals enunciated in each of the above cited works. Commentators appear to differ mostly over whether the copyright misuse doctrine is necessary or sufficient to protect the public's rights.

---

95. *Id.* at 525. Clickwrap agreements are similar to shrinkwrap agreements except that the user, instead of agreeing to the contract by opening the packaging, agrees to the contract by clicking an "I accept" box while installing the software on a computer. *Id.* at 507.

96. *Id.* at 525.

97. *See, e.g., id.* at 521.

98. *Id.*

99. *See id.* at 495-501 (arguing that copyright misuse is needed to stem the overreaching of copyright owners in clickwrap and shrinkwrap contracts).

100. *See id.* at 535.

101. Judge, *supra* note 8, at 945-47.

102. *Id.* at 930.

103. *Id.* at 931. Judge concludes that "[a]t least for now, the addition of a single policy goal (the values embodied in the First Amendment) and two guidelines to protect it (fair use and the idea/expression dichotomy) are all that is needed to address essentially all of the problematic behavior discussed in the case law thus far." *Id.* at 935.

It is interesting to read Loren's comments regarding the enforceability of some clickwrap provisions and the resulting preference of copyright holders to sue under copyright infringement instead of under breach of contract.[104] In fact, the most recent cases may tend to embolden copyright holders in their contracting behavior and encourage suits under breach of contract, where the copyright misuse defense is mute.[105]

III. FREEDOM OF CONTRACT

A. When the Courts Do Not Apply Copyright Misuse

The circuit court cases discussed in Part II(a) each analyze a contract provision (a license) for indications of copyright misuse.[106] In deciding these cases, none of the courts discussed the freedom that entities have to contract with one another.[107] The commentaries reviewing the copyright misuse cases, however, do discuss the balance between copyright misuse and freedom of contract.[108] The court opinions presented in this section take a freedom of

---

104. *See* Loren, *supra* note 7, at 512.

> [T]he possibility of content owners asserting a claim for breach of these overreaching clauses may actually be quite slim, making determinations on the enforceable nature of such clauses or the preemptive effect of the Copyright Act on such clauses exceedingly rare. . . . [I]t may be the desire on the part of the content owner to preserve the *in terrorem* effect these agreements may have. Even when an individual user has breached one of these clauses, the content owner may perceive the better strategy to ignore the breach, or at least not file a lawsuit claiming breach, rather than risk a court ruling that the clause is invalid.

Id.

105. *See* Altera Corp. v. Clear Logic, Inc., 424 F.3d 1079 (2005) (upholding standard license agreement against reverse engineering); Davidson & Assocs. v. Jung, 422 F.3d 630, 632 (8th Cir. 2005) (upholding clickwrap contract provisions against reverse engineering).

106. *See* cases cited *supra* note 12.

107. *See* cases cited *supra* note 12. *But see* Assessment Techs. of Wis., LLC v. WIREdata, Inc., 350 F.3d 640, 645 (7th Cir. 2003) (discussing the different rights under contract versus copyright infringement but finding this irrelevant because the defendant was not a party to the contracts and no breach of contract claim was brought). Arguably, the court in *Video Pipeline* considered the freedom of contract doctrine, although the opinion did not use those terms explicitly. Video Pipeline, Inc. v. Buena Vista Home Entm't, Inc., 342 F.3d 191, 206 (3d Cir. 2003).

108. *See, e.g.*, Frischmann & Moylan, *supra* note 2, at 914 ("Licenses remain a central form of legal protection, complementing or possibly even substituting for intellectual property rights."); Judge *supra* note 8, at 905 ("[T]he proposed modifications to copyright misuse [are not] intended to deter parties from contracting around the default allocation of rights embodied in the Copyright Act . . . . Rather, the proposed regime aims to force copyright holders to choose between respecting certain elements of copyright law deemed particularly important to copyright policy or forgoing their right to seek injunctive relief under the Copyright Act."); Judge *supra* note 8, at 927 ("[O]ne way of approaching the question of when attempts by copyright holders to expand their control outside the formal bounds of copyright law should be deemed misuse would be to provide copyright holders with great discretion in licensing practices but identify certain bounds as so central to copyright policy that if transgressed, the

contract approach and do not discuss the concept of copyright misuse.[109] Circuit courts that have not expressly accepted the copyright misuse doctrine have generally taken a freedom of contract approach.[110] The Ninth Circuit, despite having adopted the copyright misuse doctrine, decided a recent case, *Altera Corp. v. Clear Logic, Inc.*,[111] strictly from a contract perspective.[112]

1. *ProCD, Inc. v. Zeidenberg* (7th Cir. 1996)

Shrinkwrap licenses have become ubiquitous.[113] One of the first circuit courts to consider the enforceability of such licenses was the Seventh Circuit in *ProCD, Inc. v. Zeidenberg*.[114] "ProCD offers software and data for two prices: one for personal use, a higher price for commercial use."[115] ProCD "turned to the institution of contract" to sell a version of its program, with access to over 3,000 telephone directories, for consumer use only.[116] Zeidenberg bought a consumer version of the package, ignored the licensing agreement, set up a web-based business, and used ProCD's product to sell listings; commercial customers could get listings more cheaply from Zeidenberg than if they bought the commercial version of ProCD's system.[117] The Seventh Circuit held that the shrinkwrap license was an enforceable contract under Wisconsin state law.[118] The court also held that the Copyright Act did not preempt the license agreement.[119] Section 301(a) "prevents states from substituting their own regulatory systems for those of the national government . . . . [It] does not . . . interfere with private transactions in intellectual property, so it does not prevent states from respecting those transactions."[120] Furthermore, the court noted that ProCD's practice "may even make information more readily available by reducing the price ProCD charges to consumer buyers."[121]

---

copyright would become unenforceable (at least as a property right).").

109. *See e.g.*, *Altera*, 424 F.3d at 1089-90; *Davidson*, 422 F.3d at 638-39; Bowers v. Baystate Techs., Inc., 320 F.3d 1317, 1323-26 (Fed. Cir. 2003) ("Courts respect freedom of contract and do not lightly set aside freely-entered agreements.").

110. *See Davidson*, 422 F.3d 630; *Bowers*, 320 F.3d 1317; ProCD, Inc. v. Zeidenberg, 86 F.3d 1447 (7th Cir. 1996).

111. 424 F.3d 1079 (9th Cir. 2005).

112. *See id.* at 1089.

113. *See* Loren, *supra* note 7, at 535.

114. 86 F.3d 1447 (7th Cir. 1996).

115. *Id.* at 1454.

116. *Id.* at 1449-50.

117. *Id.* at 1450.

118. *Id.* at 1450-53.

119. *Id.* at 1455.

120. *Id.*

121. *Id.* Making works more widely available is consistent with copyright policy goals.

2. *Bowers v. Baystate Technologies, Inc.* (Fed. Cir. 2003)

In *Bowers v. Baystate Technologies, Inc.*,[122] the Federal Circuit upheld a contract that prohibited reverse engineering of the purchased software under a state law contract analysis.[123] Harold L. Bowers developed software to enhance already existing computer-aided design (CAD) software.[124] Mr. Bowers' competitor, Baystate, "contend[ed] that the Copyright Act preempts the prohibition of reverse engineering embodied in Mr. Bowers' shrink-wrap license agreements."[125] The Federal Circuit reviewed a case in which the First Circuit held that a state law trade secret claim was not preempted by the Copyright Act.[126] The court also noted that "most courts . . . have found that the Copyright Act does not preempt contractual constraints on copyrighted articles."[127] While reaffirming its holding in *Atari Games Corp. v. Nintendo of America, Inc.*[128] that "'reverse engineering object code to discern the unprotectable ideas in a computer program is a fair use,'"[129] the Federal Circuit noted that "[t]he First Circuit recognizes contractual waiver of affirmative defenses and statutory rights."[130] The court concluded that "the First Circuit would find that private parties are free to contractually forego the limited ability to reverse engineer a software product under the exemptions of the Copyright Act."[131] There is no mention of the copyright misuse doctrine in the *Bowers* opinion.[132] Judge Dyk's dissent argued that although parties may properly negotiate away their statutory rights, the shrinkwrap licence enveloping the CAD software was a contract of adhesion; agreeing to its terms did not involve true negotiation.[133]

---

122. 320 F.3d 1317 (Fed. Cir. 2003).

123. *Id.* at 1320-22. The Federal Circuit had jurisdiction because the controversy included patent as well as copyright and contract claims. *Id.* at 1322.

124. *Id.* at 1320-21.

125. *Id.* at 1323.

126. *Id.* at 1324. The Federal Circuit follows "the law of the circuit from which the appeal is taken" when deciding issues beyond its unique subject matter jurisdiction. *Id.* at 1322-23. Thus, the Federal Circuit in *Bowers* applied First Circuit law. *Id.* at 1322.

127. *Id.* at 1324.

128. 975 F.2d 832 (Fed. Cir. 1992).

129. *Bowers*, 320 F.3d at 1325 (quoting *Atari*, 975 F.2d at 843). *See also* Andrews, *supra* note 26, at 986-88 (discussing fair use as a defense to the reverse engineering of software).

130. *Bowers*, 320 F.3d at 1325.

131. *Id.* at 1325-26.

132. *Id.* at 1323-28. *See also* Andrews, *supra* note 26, at 1006 (stating that "the copyright misuse defense was not even raised in the Federal Circuit"). Andrews speculates that Baystate did not raise the copyright misuse defense because the Federal Circuit had intimated that it would apply an "antitrust approach, directed by a rule of reason analysis" to a copyright misuse defense, and that Baystate could not prove such a standard. *Id.* The lack of a copyright misuse defense could simply be due to the fact that the First Circuit had not recognized the copyright misuse defense and the Federal Circuit would apply First Circuit law to the question. *See, e.g.*, *Bowers*, 320 F.3d at 1322.

133. *Bowers*, 320 F.3d 1335-37. *See also* Judge, *supra* note 8, at 940:

3. *Davidson & Associates v. Jung* (8th Cir. 2005)

In *Davidson & Associates v. Jung*,[134] the Eighth Circuit Court upheld a lower court's summary judgment against the defendants on breach of contract claims.[135] The court also upheld a summary judgment dismissing the defendant's claim that the contracts were unenforceable.[136] The contract claims at issue involved licensing agreements with clauses that prohibited reverse engineering.[137] The two license agreements covered computer video games and the use of Battle.net, an internet site that facilitated multi-player game playing.[138] Blizzard developed "'Battle.net,' a 24-hour online-gaming service available exclusively to purchasers of its computer games."[139] Battle.net uses a "secret handshake" to allow only authenticated versions of Blizzard games to be played on its multi-game platform.[140] When installing a Blizzard game, a user must agree to an End User License Agreement (EULA), which applies to the game itself, and to a Terms of Use Agreement (TOU) which applies to participation in activities on Battle.net.[141] Both agreements prohibit reverse engineering.[142] After Jung and other Blizzard game subscribers began experiencing problems with the Battle.net server, they used reverse engineering to develop an open-source website, bnetd.org, which they designed to emulate the experience of playing Blizzard games on Battle.net.[143]

The Eighth Circuit considered a small sub-set of the original claims because many of the original claims, including claims of copyright infringement, were previously settled with a consent decree by the district court.[144] The court concluded that the contract provisions against reverse engineering were not preempted by the Copyright Act.[145] The court dismissed the defendants' reliance on *Vault Corp. v. Quiad Software Ltd.*,[146] and held that "[u]nlike in *Vault*, the state law at issue here [rights of private citizens to

---

I suggest that a copyright holder should be able to overcome a presumption of misuse arising from terms in a licensing agreement if the copyright holder can show that the terms were not included in any standardized license agreement it attached to his copyrighted work and that the copyright holder had licensed or sold the work free of the offensive provisions.

Id.

134. 422 F.3d 630 (8th Cir. 2005).

135. *Id.* at 632.

136. *Id.*

137. *Id.* at 634 n.4.

138. *Id.* at 633.

139. *Id.*

140. *Id.* at 633-35.

141. *Id.* at 634-35.

142. *Id.* at 634-35 nn. 4-5.

143. *Id.* at 635-36. Jung and his colleagues called their group the "bnetd project," and the open-source website they developed was www.bnetd.org. *Id.* at 635.

144. *Id.* at 637.

145. *Id.* at 638-39.

146. 847 F.2d 255, 268-70 (5th Cir. 1988).

contract] neither conflicts with the interoperability exception under 17 U.S.C. § 1201(f) nor restricts rights given under federal law."[147] The court favorably cited *Bowers v. Baystate* in holding that private parties may contract around default copyright limitations.[148] The court concluded that "[b]y signing the TOUs and EULAs, Appellants expressly relinquished their rights to reverse engineer."[149]

The district court had addressed the defendants' claims that (1) the "clickwrap" licenses are unenforceable contracts of adhesion, (2) the "EULA and TOU . . . are unenforceable because they prohibit the fair use of the Blizzard software," and (3) the "EULA and TOU terms constitute a misuse of Blizzard's copyright in the software."[150] Distinguishing the key copyright misuse cases upon the facts, the district court did not find copyright misuse.[151]

The language used does not prevent defendants from competing with Blizzard by prohibiting them or their employees from developing video game software as in *Lasercomb*, or require defendants to use or buy only Blizzard games similar to the AMA's prohibition on use of a competitor's coding system in *Practice Management*. The parties can terminate the licenses at any time.[152]

The district court was also reluctant to allow "the copyright misuse defense as a defense to a contract claim," as the plaintiff's copyright infringement claim was settled by the parties and dismissed in the consent decree.[153] Preemption is the only issue that the Eighth Circuit discussed in affirming the district court's first three holdings.[154] On summary judgment, the district court held that "(1) Blizzard's software end-user license and terms of usage agreements were enforceable contracts; (2) Appellants waived any 'fair use' defense; [and] (3) the agreements did not constitute misuse of copyright . . . ."[155]

4. *Altera Corp. v. Clear Logic, Inc.* (9th Cir. 2005)

The Ninth Circuit's September 2005 case, *Altera Corp. v. Clear Logic, Inc.*,[156] is the most recent sidestep of the copyright misuse doctrine. The computer software licensing agreement at issue in *Altera* stated, in relevant part, "that customers may 'use the Licensed Programs for the sole purpose of programming logic devices manufactured by ALTERA and sold by ALTERA

147. *Davidson*, 422 F.3d at 639.

148. *Id.* ("[P]rivate parties are free to contractually forego the limited ability to reverse engineer a software product under the exemptions of the Copyright Act." (alteration in original) (quoting Bowers v. Baystate Techs., Inc., 320 F.3d 1317, 1325-26 (Fed. Cir. 2003))).

149. *Id.*

150. Davidson & Assocs., Inc. v. Internet Gateway, 334 F. Supp. 2d 1164, 1176-81 (E.D. Mo. 2004), *aff'd sub nom.* Davidson & Assocs. v. Jung, 422 F.3d 630 (8th Cir. 2005).

151. *Internet Gateway*, 334 F. Supp. 2d at 1181-83.

152. *Id.* at 1182.

153. *Id.* at 1182-83.

154. *Davidson*, 422. F.3d at 638-39.

155. *Id.* at 632.

156. 424 F.3d 1079 (9th Cir. 2005).

or its authorized distributors.'"[157] Altera alleged infringement claims against Clear Logic for copying its semiconductor chips under the Semiconductor Chip Protection Act of 1984.[158] Additionally, Altera alleged "state law claims against Clear Logic for intentionally inducing Altera's customers to breach their software license agreements with Altera and also for intentional interference with those contractual relations."[159] The jury found in favor of Altera on all grounds, and the district court awarded Altera "$30.6 million in damages, $5.4 million in prejudgment interest and $394,791.68 in costs."[160] The Ninth Circuit Court affirmed both the monetary award and a permanent injunction against Clear Logic on all wrongful conduct.[161]

"Altera and Clear Logic are competitors in the semiconductor industry."[162] Each provides semiconductor chips to customers for inclusion in the customers' products.[163] Altera's semiconductor chips are programmable logic devices (PLD) whereas Clear Logic's chips are not programmable; Clear Logic's chips, Application-Specific Integrated Circuits (ASICs), perform one function.[164] Altera may work with a given customer for months to produce a PLD that performs exactly to the customer's specifications.[165] Altera sells the PLDs and licenses the required programming software.[166] When Altera's customers program an Altera device, "a file called a bitstream is generated."[167] "Clear Logic asks [its] customers to send [it] the bitstream" from the Altera device, and "Clear Logic uses the bitstream to create an ASIC for the customer."[168] By starting with the bitsream that Altera's customers generate using Altera's software, Clear Logic is able to obtain "a turnaround time of just a few weeks, and rarely produces an incompatible chip."[169]

After addressing the semiconductor chip infringement allegation, the Ninth Circuit rejected Clear Logic's contention that the state law contract claims are preempted by the Copyright Act.[170] The circuit court agreed with the district court that the "sole use" provision in the computer software license is an "extra element" which saves the state law claim from preemption.[171] The Ninth Circuit noted that "[m]ost courts have held that the Copyright Act does *not*

157. *Id.* at 1082.
158. *Id.* at 1081.
159. *Id.*
160. *Id.*
161. *Id.*
162. *Id.*
163. *Id.* at 1081 n.1.
164. *Id.* at 1081-82.
165. *Id.* at 1081.
166. *Id.* at 1081-82.
167. *Id.* at 1082.
168. *Id.*
169. *Id.*
170. *Id.* at 1089.
171. *Id.*

preempt the enforcement of contractual rights."[172] The court noted a direct analogy between ProCD's licensing constraint on the use of telephone listings and Altera's licensing constraint on the use of the bitstream.[173] The court held that "[a] state law tort claim concerning the unauthorized use of the software's end-product is not within the rights protected by the federal Copyright Act," and thus not preempted.[174]

Next, the Ninth Circuit addressed Clear Logic's copyright misuse argument that "Altera's license agreements should not be enforced because they operate as 'illegal tying arrangements because Altera is using its copyright in its programming software to control competition in an area outside the copyright, i.e. Altera's chip hardware.'"[175] The Ninth Circuit briefly reviewed its adoption of copyright misuse in *Practice Management Information Corp. v. American Medical Ass'n* as a "'*defense* to copyright infringement.'"[176] The Court went on to state the following:

> When copyright misuse applies, we do not allow enforcement of the copyright for the period of misuse. Because the remedy for copyright misuse is equitable, it makes little sense to allow Clear Logic to proceed on an independent claim for copyright misuse when there has been no allegation of copyright infringement. We have already rejected Clear Logic's copyright preemption argument. We cannot now void the license agreements under the pretext of refusing to enforce a copyright that has not been asserted. Copyright misuse is not a defense to the state law claims asserted by Altera.[177]

The Ninth Circuit quoted *Alcatel USA, Inc. v. DGI Technologies., Inc. (DSC II)*[178] for the proposition that copyright misuse "'forbids the use of the [copyright] to secure an exclusive right or limited monopoly not granted by the [Copyright] Office and which is contrary to public policy to grant.'"[179] The Ninth Circuit, however, ignored any similarities between the business models of DSC and Altera while readily accepting the differing nature of the claims in *DSC II* and *Altera*.[180]

---

172. *Id.*

173. *Id.*

174. *Id.* at 1090.

175. *Id.*

176. *Id.* (quoting Practice Mgmt. Info. Corp. v. Am. Med. Ass'n, 121 F.3d 516, 520 (9th Cir. 1997)) (emphasis in original).

177. *Id.* (internal citations omitted).

178. 166 F.3d 772 (5th Cir. 1999).

179. *Altera*, 424 F.3d at 1090 (quoting *DSC II*, 166 F.3d at 792).

180. *Id.*

B. *The Concern with Freedom of Contract in Copyright Analysis*

Academic commentators have almost universally criticized cases like *ProCD* and *Bowers*.[181] As David A. Rice stated: "Judge Easterbrook's *ProCD* opinion is often criticized by legal scholars but, as it was in *Bowers*, often embraced by other courts without reflection."[182] The Federal Circuit's reasoning in *Bowers* is cited and followed by the Eighth and Ninth Circuit Courts in *Davidson* and *Altera*.[183] Given the apparent influence of this case on the analytical approach to these more recent cases (i.e., freedom of contract versus copyright misuse), the *Bowers* case deserves further analysis. In *Copyright and Contract: Preemption after Bowers v. Baystate*,[184] Rice strongly disagrees with the reasoning and the outcome of the Federal Circuit in deciding *Bowers v. Baystate Technologies, Inc.*[185] According to Rice, the court's reasoning (1) misapplies First Circuit law, (2) uses "highly selective, superficial and even misleading" case law analysis, (3) errs in holding "that state contract law may be employed to circumvent one provision of the federal Copyright Act that explicitly narrows the statute-created exclusive rights of a copyright owner set forth in another section of the Act," and (4) "pa[ys] no heed to the two ultimate consequences of its decision"–expanded copyright scope and impact on state trade secret law.[186] Rice differentiates between the use of contracts to facilitate the exercise of an author's copyright grant, which typically and appropriately is not preempted by the Copyright Act, and the use of contracts to expand the scope of that grant.[187] Instead of using the "judicially-devised and non-statutory [extra element test],"[188] for preemption, "federal courts ought to consider whether the same conduct or acts establish both state and federal law liability when making their determination of the

---

181. *See, e.g.*, Loren, *supra* note 7, at 507 (citing several instances of "widespread scholarly criticism of the [ProCD] decision"); David A. Rice, *Copyright and Contract: Preemption after Bowers v. Baystate*, 9 Roger Williams U. L. Rev. 595, 595-644 (2004) (strongly criticizing the decision in *Bowers*).

182. Rice, *supra* note 181, at 638.

183. *See Altera*, 424 F.3d at 1089; *Davidson*, 422 F.3d at 639.

184. Rice, *supra* note 181.

185. 320 F.3d 1317 (Fed. Cir. 2003).

186. Rice, *supra* note 181, at 598-99.

187. *See* Rice, *supra* note 181, at 615-16.

> State contract law is the principal means for exercise and realization of the economic benefits of copyright owner rights created by the Copyright Act . . . . A sizeable majority of the decisions have held that the term at issue was not preempted. This is quite as one would expect. Contract is the means for copyright owners' exploitation of limited property rights through transfer of exclusive or non-exclusive rights.

Id. at 615-616 (citation omitted).

Rice goes on to argue that "[i]t defies imagination that Congress intended, but did not state, that general contract law and the Uniform Commercial Code could be used to enforce standard form, market-wide 'private legilsation' to repeal [§] 107 and, with that, effectively rewrite [§] 102(b)." *Id.* at 639.

188. *Id.* at 616.

equivalency of rights created by state law and federal copyright law."[189] Rice contends that preemption is compatible with the copyright misuse doctrine and its "necessity of case-by-case litigation."[190] Finally, Rice advocates overturning *Bowers*, either judicially or legislatively.[191]

> Whether the correction [of *Bowers v. Baystate*] emanates from the Court or Congress, its message will be that federal courts must refrain from complicity with copyright owners' presumptuously preemptive misuse of state contract law to substantially enlarge their substantial but limited statutory rights by diminishing or zeroing out intended public benefits of copyright.[192]

IV. CONTRACT-BASED COPYRIGHT MISUSE VERSUS FREEDOM OF CONTRACT

The cases discussed in Part II(a) and Part III(a) show a divergent framework for assessing the behaviors and rights of a copyright licensor that depends on (1) the court's view of the appropriateness with which a copyright holder may contractually extend his rights and (2) whether a copyright infringement claim is before the court. This divergence is especially clear in *Altera* and *DSC II*, when the courts applied completely different reasoning to similar issues involving companies with similar business models. Those cases are compared in detail below.

A. Altera and DSC II Compared

As the defendant in the *Altera* case, Clear Logic argued that "Altera's license agreements should not be enforced because they operate as 'illegal tying arrangements because Altera is using its copyright in its programming software to control competition in an area outside the copyright, i.e. Altera's chip hardware.'"[193] Not surprisingly, Clear Logic's argument parallels the Fifth Circuit's holding in *Alcatel USA, Inc. v. DGI Technologies, Inc. (DSC II)*.[194] Altera customers were licensed to use Altera's copyrighted software only to program Altera's chips,[195] and DSC's customers were licensed to use DSC's copyrighted software only to operate DSC's phone switches.[196] The *DSC II* court held that the district court abused its discretion in ignoring the jury's finding of copyright misuse,[197] stating that "[a] reasonable juror could

189. *Id.* at 621.

190. *Id.* at. 640.

191. *Id.* at 644.

192. *Id.* Unfortunately, from Prof. Rice's perspective, the two recent circuit court decisions follow the *Bowers* Court. *See* Altera Corp. v. Clear Logic, Inc., 424 F.3d 1079, 1089 (9th Cir. 2005); Davidson & Assocs. v. Jung, 422 F.3d 630, 639 (8th Cir. 2005).

193. *Altera*, 424 F.3d at 1090.

194. *See generally DSC II*, 166 F.3d 772 (5th Cir. 1999).

195. *Altera*, 424 F.3d at 1082.

196. *DSC II*, 166 F.3d at 777.

197. *Id.* at 794.

conclude, based upon the licensing agreement, that 'DSC has used its copyrights to indirectly gain commercial control over products DSC does not have copyrighted,' namely, its microprocessor cards."[198]

Given that the parallel licensing agreements in the two cases allowed computer software use only in conjunction with the vendor's hardware, why were there such different outcomes? The *DSC II* case in the Fifth Circuit resulted in no injunction based upon copyright infringement and a finding of copyright misuse against the licensor,[199] while the recent *Altera* case in the Ninth Circuit resulted in a permanent injunction and over thirty-five million dollars in damages for the licensor.[200] Having previously adopted the copyright misuse doctrine, the Ninth Circuit's *Altera* opinion quotes the *DSC II* Court's description of "copyright misuse as an 'unclean hands defense' which 'forbids the use of the [copyright] to secure an exclusive right or limited monopoly not granted by the [Copyright] Office and which is contrary to public policy to grant.'"[201] The court went on to conclude, however, that "[c]opyright misuse is not a defense to the state law claims asserted by Altera."[202]

Following the Ninth Circuit's logic, had DSC alleged state law contract claims instead of copyright infringement claims, the outcome in *DSC II* and *Altera* could have been similar. In other words, if in 1995 DSC had filed a supplemental complaint asserting "state law claims against [DGI] for intentionally inducing [DSC's] customers to breach their . . . license agreements with [DSC] and also for intentional interference with those contractual relations"[203] instead of "asserting direct and indirect copyright infringement claims,"[204] the outcome in that case could have been similar to the later *Altera* holding. With no copyright infringement allegations, the court would not have entertained a copyright misuse defense, and DGI would have been found liable for the state law contract claims.

B. Sidestepping Copyright Misuse with State Law Contract Claims

A direct comparison between *Altera* and *DSC II* is irresistible given the

198. *Id.* at 793.

199. *Id.* at 799. The case was remanded to modify the injunction as the only basis that was upheld was trade secret misappropriation. *Id.*

200. *Altera*, 424 F.3d at 1081. The damages and injunction are based upon all claims; infringement under SCPA and "state law claims . . . for intentionally inducing Altera's customers to breach their software license agreements with Altera and also for intentional interference with those contractual relations." *Id.*

201. *Id.* at 1090 (quoting *DSC II*, 166 F.3d at 792 (5th Cir. 1999)) (alterations in original) (internal citations omitted).

202. *Id.* Altera may have used a breach of contract argument to avoid the copyright misuse defense that hurt Alcatel. Further, Altera may not have had a straightforward copyright infringement claim because Clear Logic did not use a copy of their copyrighted program.

203. *Id.* at 1081 (alterations are a modification of the claims Altera brought against Clear Logic).

204. *DSC II*, 166 F.3d at 779.

different analysis and opposite impact of the courts' decisions. Companies in both cases had licensing provisions designed to use copyrighted software to thwart competition in hardware. One company's licensing provisions were deemed enforceable, while the other company was turned away from the court. An important question stemming from the comparison is whether a finding of copyright misuse can simply be sidestepped by asserting a state law interference with contracts claim instead of a copyright infringement claim. If the answer, based upon the most recent cases, is yes, then what, if anything, should be done about it?

The question, however, goes beyond the application of copyright misuse. The circuit courts that applied copyright misuse either explicitly or implicitly considered whether the licensing agreement inappropriately expanded the copyright owner's rights. The circuit courts that used a freedom of contract Copyright Act preemption analysis (with the exception of *ProCD*[205]), however, did not appear to consider the implications for Copyright policy if the restrictive license terms were broadly applied. Rather, those courts seemed to simply accept the idea that "'most courts to examine this issue have found that the Copyright Act does not preempt contractual constraints on copyrighted articles.'"[206] Thus, courts approaching the analysis from the freedom of contract perspective enforced contracts in *ProCD* (individual, non-commercial use), *Bowers v. Baystate* (no reverse engineering), *Davidson & Associates v. Jung* (no reverse engineering), and *Altera v. Clear Logic* (use only with specific hardware). On the other hand courts considering misuse did not enforce copyright or contract claims in *Lasercomb* (no independent creation), *DSC I* and *DSC II* (use only with specific hardware), or *Practice Management* (use only with specified system).[207]

Video Pipeline is a circuit court case in which the court considered the overall goals of copyright and decided that the contract provisions, on balance and as applied to the particular case, help to enhance those goals.[208] Other

---

205. ProCd, Inc. v. Zeidenberg, 86 F.3d 1447, 1455 (7th Cir. 1996). Although Judge Easterbrook considers the impact of the license agreement on "mak[ing] information more readily available," he concludes that "whether a particular license is generous or restrictive, a simple two-party contract is not" preempted by the Copyright Act. *Id.*

206. *Altera*, 424 F.3d at 1089 (quoting Bowers v. Baystate Techs., Inc. 320 F.3d 1317, 1324-25 (Fed. Cir. 2003)); *see* Rice, *supra* note 181, at 638 (stating that "Judge Easterbrook's *ProCD* opinion is often criticized by legal scholars but, as it was in *Bowers*, often embraced by other courts without reflection").

207. *ProCD*, 86 F.3d at 1449-50; *Bowers*, 320 F.3d at 1322; Davidson & Assocs. v. Jung, 422 F.3d 630, 634 nn.4-5 (8th Cir. 2005); *Altera*, 424 F.3d at 1086-93.

208. Video Pipeline, Inc. v. Buena Vista Home Entm't, Inc., 342 F.3d 191, 206 (3d Cir. 2003) (considering that disallowing Disney any control over other content on websites where it licensed use of its trailers "would likely decrease the public's access to Disney's works because it might as a result refuse to license at all online display of its works"); *see also ProCd*, 86 F.3d at 1455 ("Enforcement of the shrinkwrap license may even make information more readily available, by reducing the price ProCd charges to consumer buyers."). Judge Easterbrook did not find a shrinkwrap contract's impact on copyright goals dispositive to the enforcement of such contracts. *Id.* He concluded that "whether a particular license is generous or restrictive, a

misuse perspective cases strongly lean toward the conclusion that increasing the rights granted under the limited monopoly of the Copyright Act is misuse, while the contract cases appear to be based upon a fiction of private-party, individually negotiated licenses.

It is challenging to balance the rights of individuals to contract and the public policy promoted through copyright. It becomes even more difficult when courts narrow their focus and do not thoughtfully consider both contract and copyright policy. Proper consideration of both policies requires a fact-specific balancing test, which may be as complex as the fair use test.

C. A Proposed Balancing Test and Example Applications

The *Bowers* court stated that "most courts to examine this issue have found that the Copyright Act does not preempt contractual constraints on copyrighted articles."[209] Contracts are the business means by which copyright holders exercise their legitimate rights; it is not surprising that most contracts would not be preempted.[210] If the contract merely facilitates the author's exercise of legitimate rights under the Copyright Act and does not otherwise impact the defendant's statutory rights, then there is no need for further analysis. The contract is enforceable and not preempted by the Copyright Act. If a contract further expands the copyright holder's rights, then a balancing test may be necessary in order to determine the contract's enforceability.[211]

The blunt instrument of copyright misuse, which renders a copyright unenforceable, is overly punitive under many circumstances when an author may have overreached.[212] Also, although courts readily note that authors may return to court when they have stopped the practice and purged the results,[213] I have not found any example of what constitutes such a purging. While courts have disagreed as to whether a defendant must have clean hands to invoke the copyright misuse defense,[214] when misuse is found, the result has been that the

---

simple two-party contract is not 'equivalent to any of the exclusive rights within the general scope of copyright' and therefore may be enforced." *Id.*

209. *Bowers*, 320 F.3d at 1324.

210. *See* Rice, *supra* note 181, at 616 ("Contract is the means for copyright owners' exploitation of limited property rights through transfer of exclusive or non-exclusive rights.").

211. Although this Comment proposes a balancing test to determine if a contract is enforceable, the criteria prompting the need for the test are equivalent to Loren's criteria for a presumption of copyright misuse. *See Loren, supra* note 7, at 501 ("propos[ing] that a rebuttable presumption of misuse should be recognized anytime copyright owners seek, by contract, to avoid the express statutory limitations on their rights").

212. *See, e.g.*, Judge, *supra* note 8, at 950 ("The incongruity of such a remedy [as unenforceability] with violations like . . . using a standardized license that effectively expands the scope of one's copyright to cover the idea that it expresses . . . may be one of the primary factors preventing misuse from being invoked . . . where it could apply.").

213. *See, e.g.*, Lasercomb Am., Inc. v. Reynolds, 911 F.2d 970, 979 n.22 (4th Cir. 1990).

214. *Compare* Alcatel USA, Inc. v. DGI Techs., Inc. 166 F.3d 772, 794, 798-99 (5th Cir. 1999) (holding that the defendant could invoke the copyright misuse defense despite its own "very dirty mitts"), *with* Atari Games Corp. v. Nintendo of Am. Inc., 975 F.2d 832, 846 (Fed.

defendant-infringer can completely escape penalty for copyright infringement (and presumably is free to continue infringing until the plaintiff has discontinued and purged the results).[215] Furthermore, copyright misuse has only been recognized as a defense to a copyright infringement claim; it can be sidestepped by bringing a breach of contract claim instead of a copyright infringement claim.[216] Given these subtleties, along with changing technologies, a balancing test to determine if the contract provision is enforceable may be useful.[217] Under such a balancing test, if the contract, as implemented, offends the overall goals of copyright, the contract is unenforceable and preempted by the Copyright Act. Application of the test would allow courts to consider both the plaintiff's and the defendant's actions in coming to a final conclusion, whereas fair use focuses solely on the defendant's actions and copyright misuse focuses solely on the plaintiff's actions. Such a test will lead to more consistent analysis of cases involving contracts that expand a copyright holder's rights.

1. The Proposed Test

The proposed balancing test asks that courts consider the contract, the circumstances in which the parties entered into the contract, and the contract's practical public impact. Based upon these factors, a court would decide whether the contract is enforceable. Just as in a fair use analysis, the factors may not "be treated in isolation, one from another[; a]ll are to be explored, and the results weighed together, in light of the purposes of copyright."[218] An explanation of the three factors and examples of their application to key cases is given below.

The exceptions to the Copyright owner's exclusive rights—that is, the rights that the rest of the world has in relation to the creative work—are alienable rights for individuals.[219] An individual is absolutely free to negotiate a deal that

---

Cir. 1992) (holding that "Atari's unclean hands prevent it from invoking equity" of the copyright misuse defense).

215. *See, e.g.*, *Lasercomb*, 911 F.2d at 980-81.

216. *See, e.g.*, Altera Corp. v. Clear Logic, Inc. 424 F.3d 1079, 1090 (9th Cir. 2005).

217. The proposed test does not give a bright line result as to which license agreements will be acceptable and which will not. Others advocate a more bright line approach, such as a few discrete sets of circumstances that would constitute copyright misuse per se. *See, e.g.*, Frischmann & Moylan, *supra* note 2, at 865.

218. Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 578 (1994).

219. For criticism of the alienability of an individual's rights with respect to a copyrighted work, see Loren, *supra* note 7, at 525-530.

> It is undeniable that the "rights" of users [of copyrighted works] have value. However, the value of most of the limitations on copyright owners [sic] rights are not values that are entirely captured by a particular user of a copyrighted work. Many of the uses permitted by the limits on copyright owners' rights have positive externalities that are not internalized by users. . . . [I]f these limitations on copyright owners [sic] rights are seen as alienable rights belonging to users, it is likely that users will undervalue those rights, and thus [sic] contract them away for little or no additional consideration.

includes, for example, his agreement not to reverse engineer a piece of software. If the contract limits a statutory right under the Copyright Act, the first factor will weigh in favor of the defendant-infringer. In *Bowers v. Baystate*, for example, the restriction against reverse engineering limited Baystate's fair use rights under the Copyright Act.[220] Likewise, the restriction requiring the use of purchased software only with the licensor's hardware in *DSC II* and *Altera* limits a competitor's ability to reverse engineer the software.[221] The first factor would weigh in favor of the defendant-infringer in each of these cases.

The second factor looks at the level of autonomy involved in contracting, with contracts of adhesion at one extreme and truly individually negotiated contracts at the other.[222] A contract of adhesion that limits a defendant-infringer's rights under the Copyright Act would weigh against the plaintiff-copyright holder. A negotiated waiver of the rights would weigh in favor of the plaintiff-copyright holder. In *Practice Management*, for example, the true uniqueness of the licensing agreement would weigh in favor of the plaintiff-copyright holder.[223]

The third factor to be considered is the protection of the public's interest in the copyrighted works. Do the contract provisions increase overall availability of the copyrighted work to the public? Can the general public find alternative means to avoid the restrictions in the license? Is the contract provision inherently more valuable to the public than to an individual user?[224] In *ProCD*, the license allowed increased public access to the copyrighted work.[225] In *Video Pipeline*, the public had many avenues to criticize Disney, and allowing the license agreement allowed increased web-access to Disney movie trailers.[226] Thus, the third factor would weigh in favor of the copyright holder in both of these cases. Limitations against reverse engineering of computer software, as in *Davidson & Associates v. Jung*, would weigh against the copyright holder.[227]

Id. at 530 (footnote omitted).

220. Bowers v. Baystate Techs., Inc., 320 F.3d 1317, 1323 (Fed. Cir. 2003); *see also* Atari Games Corp. v. Nintendo of Am. Inc., 975 F.2d 832, 843 (Fed. Cir. 1992) ("Thus, reverse engineering object code to discern the unprotectable ideas in a computer program is a fair use.").

221. *See Altera*, 424 F.3d at 1081-83; *DSC II*, 166 F.3d at 777-78.

222. *Cf.* Loren, *supra* note 7, at 511-12 (noting that clickwrap contracts that require agreement to access the copyrighted work "do appear to grant rights against the world"); Judge, *supra* note 8, at 939 (arguing that copyright holders should be able to overcome a presumption of misuse in the context of a truly negotiated contract agreement).

223. Practice Mgmt. Info. Corp. v. Am. Med. Ass'n, 121 F.3d 516, 517-18 (9th Cir. 1997).

224. *See* Loren, *supra* note 7, at 525-530.

225. ProCD, Inc. v. Zeidenberg, 86 F.3d 1447, 1455 (7th Cir. 1996).

226. Video Pipeline, Inc. v. Buena Vista Home Entm't, Inc., 342 F.3d 191, 206 (3d Cir. 2003).

227. Davidson & Assocs. v. Jung, 422 F.3d. 630, 634-35 nn. 4-5 (8th Cir. 2005); *see, e.g.*, Frischmann & Moylan, *supra* note 2, at 865; Rice, *supra* note 181, at 600-604 (arguing that *Bowers v. Baystate* was wrongly decided).

2. *DSC II, Altera, Davidson*

Consider *DSC II* and *Altera* under the three-factor test. Weighing the factors, it is likely that neither DSC nor Altera's license restrictions would be enforceable. The agreements expand the scope of copyright protection and appear to be contracts of adhesion. The ability to reverse engineer products in the market is highly valued in our society. With the balancing test, the analysis of DSC's and Altera's actions would be consistent; any outcome difference would now rest on the defendants' actions. In general, competitors of DSC and Altera may have a legitimate reverse engineering fair use defense to a copyright infringement claim.[228] DGI, however, engaged in trade secret misappropriation and does not have an effective fair use defense.[229] Under the proposed analysis, if DSC had sued under state law contract claims, the contract would be deemed unenforceable, but DGI would remain liable for its misappropriations. In the *Altera* case, Clear Logic would still face liability under the Semiconductor Chip Protection Act of 1984 (the "SCPA") for having copied, and not reverse engineered, Altera's semiconductor chip.[230] Applying the balancing test, Altera would not win its state law contract claims, and the case would need to be remanded to consider appropriate legal and equitable relief under the SCPA only. If Altera had pled copyright infringement, Clear Logic probably would not have a fair use claim on the software as such use would not likely be considered to be reverse engineering. Clear Logic's use of the incidental output of Altera's software is analogous to Zeidenberg's free riding use of the information from *ProCD*.[231] Under this analysis, Altera would be better off pleading copyright infringement.

Applying the balancing test to *Davidson & Associates v. Jung*, however, would result in a slightly different outcome. The test would render the blanket prohibition against reverse engineering unenforceable.[232] Because the

---

228. *See, e.g.*, Atari Games Corp. v. Nintendo of Am. Inc., 975 F.2d 832, 843 (Fed. Cir. 1992) ("Thus, reverse engineering object code to discern the unprotectable ideas in a computer program is fair use."). *But see* Andrews, *supra* note 26, at 1010 ("[I]t is unlikely that the reverse engineering activities in either *Alcatel* [*DCS II*] or *Bowers* would have passed the fair use test because of the effects of the product resulting from the reverse engineering on the potential market of the copyright holder.").

229. Alcatel USA, Inc. v. DGI Techs., Inc. (*DSC II*), 166 F.3d 772, 785 (5th Cir. 1999) (stating that "DGI obtained DSC's trade secrets through improper means").

230. Altera Corp. v. Clear Logic, Inc., 424 F.3d 1079, 1081, 1086-89 (9th Cir. 2005).

231. *See id.* at 1089; ProCD, Inc. v. Zeidenberg, 86 F.3d 1447, 1454 (7th Cir. 1996) ("Zeidenberg wants to use the data without paying the seller's price; . . .").

232. The contract expands Blizzard's copyright by disallowing reverse engineering, and it is a contract of adhesion; clearly, the first two factors go against Blizzard. *Davidson*, 422 F.3d at 634-35 nn.4-5. Who the third factor benefits may be debatable. "[T]he complete absence of enforcement of reverse engineering prohibitions would likely stifle the creative work (in software) that the copyright laws are intended to promote." Andrews, *supra* note 26, at 1007. Several commentators, however, see blanket reverse engineering restrictions as unacceptable because copyright does not protect ideas. *See, e.g.*, Frischmann & Moylan, *supra*, note 2, at 865 ("[P]ropos[ing] a per se rule against licensing restrictions upon reverse engineering that

copyright infringement claims had been settled,[233] the analysis would end. Thinking broadly about what Jung and his associates did, they likely have a fair use defense for the reverse engineering of the computer software.[234] They also, however, are likely liable as contributory infringers as they knowingly allowed unauthorized copies of Blizzard's software to play on bnetd.org.[235] The main litigation lesson is that acts impacting copyright law will be litigated under copyright law.

D. What Should Congress or the Supreme Court Do?

Congress has at least three basic choices for addressing the split in the courts and the trend toward enforcement of standardized contracts that effectively serve as "private legislation" expanding the copyright owners' rights against the world. First, Congress can do nothing, and perhaps the Supreme Court will take the lead. Second, Congress can fashion a Copyright Misuse Act. Both Lydia Pallas Loren and Kathryn Judge present good ideas on how the copyright misuse doctrine should be fashioned.[236] Prof. Loren's approach would basically confer a rebuttable presumption of copyright misuse on any attempt to deprive any member of the public of a statutory right under §§ 107-122.[237] Judge would limit the legislation to actions that are anticompetive or that affect free speech.[238] Finally, Congress could amend § 301 to make it explicitly clear that standard license agreements like the ones in *Bowers*, *Altera*, or *Davidson* are against the public policy behind our copyright laws and are preempted by the Copyright Act. Congress could still allow a truly negotiated contract between a copyright owner and a given individual that is consistent with the public's best interest.

The Supreme Court has rejected several opportunities to confront the issue of expanded rights through licensing agreements.[239] The Court also denied

---

complements antitrust-based misuse and the fair use doctrine."); Loren, *supra* note 7, at 523 ("[I]f a shrinkwrap or clickwrap clause purports to limit activity that a majority of courts have found to be fair use, [e.g., reverse engineering software,] that clause should also trigger a presumption of misuse."); Rice, *supra* note 181, at 622 ("Bowers . . . sanctions[s] use of standard form shink-wrap licenses and state contract law to redefine the scope of federal copyright, a wholly statutory creation."); Judge, *supra* note 8, at 931 ("[T]his Note argues that . . . any attempt by a copyright holder to cross the idea/expression boundary or to deter fair use of his copyrighted material [should be] deemed misuse."). Accordingly, considering all three factors, the blanket prohibition against reverse engineering would be unenforceable.

233. Davidson & Assocs., Inc. v. Internet Gateway, 334 F. Supp. 2d 1164, 1167 (E.D. Mo. 2004) *aff'd sub nom.* Davidson & Assocs. v. Jung, 422 F.3d 630 (8th Cir. 2005).

234. *Davidson*, 422 F.3d at 636 (describing the reverse engineering done by defendants).

235. *Id.* at 636 ("[Defendant] was aware that unauthorized versions of Blizzard games were played on bnetd.org.").

236. *See* Loren, *supra* note 7, at 522-35; Judge, *supra* note 8, at 930-951.

237. *See* Loren, *supra* note 7, at 522-25.

238. *See* Judge, *supra* note 8, at 930-31.

239. Video Pipeline, Inc. v. Buena Vista Home Entm't, Inc., 342 F.3d 191 (3d Cir. 2003), *cert. denied*, 540 U.S. 1178 (2004); Practice Mgmt. Info. Corp. v. Am. Med. Ass'n, 121 F.3d

certiorari petitions for *Bowers*.[240] Unfortunately, the Court was not asked to grant certiorari for *Altera* or *Davidson*,[241] cases that would have benefitted from a joint decision. If the Court continues to deny certiorari to these copyright/contract cases and Congress does not act, copyright holders are likely to be emboldened by the recent circuit court holdings and press for stronger contracts, which could result in more "private legislation."

V. CONCLUSION

The circuit courts are split as to when and how copyright owners may expand their rights beyond the limited rights granted by Congress in the Copyright Act. The courts approach the problem from two different perspectives: copyright misuse and freedom of contract. A court relying on one of these perspectives usually fails to recognize or discuss the other perspective, and results may vary drastically depending upon which approach a court takes. While the articulated balancing test is one possible solution, the law in this important area will remain unpredictable until either Congress or the Supreme Court decides to step in.

JENNIFER R. KNIGHT

516 (9th Cir. 1997), *cert. denied*, 524 U.S. 952 (1998).

240. Bowers v. Baystate Techs., Inc., 320 F.3d 1317 (Fed. Cir. 2003), *cert denied*, 539 U.S. 928 (2003).

241. A petition for a writ of certiorari must be filed "within 90 days after entry of the judgment" by a United States court of appeals. SUP. CT. R. 13.

TENNESSEE'S NEW LIMITED LIABILITY COMPANY ACT: NEW WAYS OF DOING BUSINESS

I. INTRODUCTION

The Tennessee legislature passed Public Chapter 286 on May 16, 2005 and made sweeping revisions to the operation of limited liability companies ("LLCs") in Tennessee.[1] A revision committee, appointed by the Tennessee Bar Association and consisting largely of practicing Tennessee attorneys, drafted the legislation.[2] Public Chapter 286 (the "New Act") became effective on January 1, 2006.[3] Several new and interesting features to the New Act are discussed below; however, this Comment focuses primarily on a discussion of the three management paradigms offered by the New Act.

Part II of this Comment explores the history of the limited liability company since its beginning in the United States and discusses the common features of LLCs found throughout different jurisdictions. Additionally, Part II takes a brief look at the effects federal income tax considerations have had on the development of LLCs. Finally, Part II addresses Tennessee's 1994 LLC Act and amendments and concludes with a brief introduction to the New Act.

Finally, Part III of this Comment examines each of the three LLC management paradigms under the New Act. Particular consideration is given to the organizational structure and agency relationships within each management paradigm. The fiduciary duties of the relevant parties within each management paradigm are also compared.

II. BACKGROUND

A. *A Brief History of the Limited Liability Company*

The limited liability company found its United States origins in the 1977 Wyoming LLC Act.[4] The popularity of the LLC has flourished since the IRS

---

1. Tennessee Revised Limited Liability Company Act, ch. 286, 2005 Tenn. Pub. Acts 1-109. codified at TENN. CODE ANN. §§ 48-249-101 to -249-1133 (Supp. 2005).

2. Richard R. Spore, III served as Chairman and Reporter. The remaining committee members include: Arnold B. Clapp, Van Philip East, III, B. Riney Green, Joan MacLeod Heminway, J. Nelson Irvine, Richard A. Johnson, James Howard Levine, William E. McClamroch, III, J. Timothy Morris McLemore, Warner B. Rodda, Samuel E. Stumpf, Jr., Jason G. Yarbro, and G. Michael Yopp. TENNBARU, UNDERSTANDING THE NEW LIMITED LIABILITY COMPANY ACT sched. I at 29-30 (July 29, 2005) [hereinafter TENNBARU]. Professor Joan Heminway has provided the author with a significant amount of guidance and insight regarding the development of the committee's progress and business organization law in general, without which this comment would not have been possible.

3. *Id.* at SECTION 2.

4. William J. Carney, *Limited Liability Companies: Origins and Antecedents*, 66 U. COLO. L. REV. 855, 857 (1995); *see also* Wyoming LLC Act, 1977 Wyo. Sess. Laws 537 (codified at

announced through Revenue Ruling 88-76 that an LLC may be taxed as a partnership rather than a corporation.[5] Specifically, the IRS held in Revenue Ruling 88-76 that an LLC will be treated "as a partnership for federal tax purposes unless the organization has a preponderance of the . . . corporate characteristics of continuity of life, centralization of management, limited liability, and free transferability of interests."[6]

In 1992, Delaware enacted the Delaware Limited Liability Company Act.[7] Delaware has always been a popular jurisdiction for business entity formation, and LLCs are no exception.[8] The expansive "freedom of contract" policy—one of the hallmarks of modern LLC acts across the United States[9]—was codified by the Delaware legislature along with an emphasis on the enforceability of LLC agreements.[10]

Delaware has chosen the traditional and most popular governance schemes in the operation and nomenclature of its LLC management paradigms.[11] Two management paradigms are available for LLCs organized under Delaware law: (1) *member-managed* and (2) *manager-managed*.[12] The member-managed LLC is the functional equivalent of a partnership, that is the members, like partners in a partnership, retain ownership and control over the dealings of the LLC.[13] The manager-managed LLC is the functional equivalent of a limited partnership, wherein the managers, like general partners, have control over the dealings of the LLC and the members, like limited partners, retain ownership.[14]

WYO. STAT. ANN. §§ 17-15-101 to -144 (1977)).

5. Carney, *supra* note 4, at 858.

6. Rev. Rul. 88-76, 1988-2 C.B. 360, 361.

7. Limited Liability Company Act, ch. 434, 68 Del. Laws 1329 (1992) (codified as amended at DEL. CODE ANN. tit. 6, §§ 18-101 to -1109 (1999)).

8. CARTER G. BISHOP & DANIEL S. KLEINBERGER, LIMITED LIABILITY COMPANIES: TAX AND BUSINESS LAW ¶ 14.01[2] (2005) (citing Walter C. Tuthill & Louis G. Hering, *Delaware Lawmakers Amend Alternative Entity Statutes*, 15 BNA CORP. COUNS. WKLY., Aug. 2, 2000, at 8).

9. Carney, *supra* note 4, at 877.

10. DEL. CODE ANN. tit. 6, § 18-1101(b) (1999). The Delaware Supreme Court noted the following about freedom of contract in Delaware LLCs:

> The Act can be characterized as a "flexible statute" because it generally permits members to engage in private ordering with substantial freedom of contract to govern their relationship, provided they do not contravene any mandatory provisions of the Act. Indeed, the LLC has been characterized as the "best of both worlds."

Elf Atochem N. Am., Inc. v. Jaffari, 727 A.2d 286, 290 (Del. 1999) (footnote omitted) (quoting MARTIN I. LUBAROFF & PAUL M. ALTMAN, *Delaware Limited Liability Companies*, *in* DELAWARE LAW OF CORPORATIONS & BUSINESS ORGANIZATIONS, § 20.1 ®. Franklin Balotti & Jesse A. Finkelstein eds., 1987)).

11. *See* BISHOP & KLEINBERGER, *supra* note 8, at ¶ 7.02[1][b] (citing Delaware, Florida, and California statutes).

12. *See* DEL. CODE ANN. tit. 6, § 18-402 (1999).

13. *See id.*

14. *See id.* In a manager-managed LLC, some would equate members with corporate

In 1996, the Uniform Limited Liability Company Act ("ULLCA") was drafted by the National Conference of Commissioners on Uniform State Laws; the ULLCA permits the operation of LLCs as either member-managed or manager-managed companies.[15] Like Delaware and most other states, the uniform Act encourages the "freedom of contract" policy.[16] The "freedom of contract" policy emphasized in the ULLCA has been described as follows:

> The ULLCA is a completely flexible act that combines business and tax default rules with the freedom of contract. The default rules regarding continuity of life, free transferability of interests, and centralized management are designed to ensure that an LLC organized in a state that adopts the ULLCA will be classified as a partnership for federal tax purposes. *Nearly all the act's rules, including the rule regarding limited liability, are simply default rules that may be modified in the operating agreement or in the articles of organization.* Accordingly, members have the freedom of contract to specially tailor an LLC to their specific needs, which may even include having the company taxed like a corporation for federal tax purposes.[17]

Another practical and timesaving feature of the ULLCA is the consolidation of all nonwaivable provisions into a single section of the Act.[18] Other than the seven nonwaivable provisions listed in section 103(b) of the ULLCA, "all members of a limited liability company may enter into an operating agreement... to regulate the affairs of the company and the conduct of its business, and to govern relations among the members, managers, and company."[19] At least four states have found these features attractive and have adopted the ULLCA.[20]

B. Tennessee's Limited Liability Company Act Prior to January 1, 2006

In 1994, Tennessee became the 40th state to endorse the formation of LLCs by passing the Tennessee Limited Liability Company Act (the "Prior Act").[21] With the IRS's Revenue Ruling 88-76 in mind,[22] the drafters of the Prior Act

---

shareholders and managers with corporate directors. *See, e.g.*, WILLIAM A. GREGORY, THE LAW OF AGENCY AND PARTNERSHIP 510 (3d ed. 2001).

15. UNIF. LTD. LIAB. CO. ACT § 203 cmt. (1996) [hereinafter ULLCA] ("A company will be member-managed unless it is designated as manager-managed under Section 203(a)(6).").

16. ULLCA, Prefatory Note 2 (1996) ("Only limited and specific fundamental matters may not be altered by private agreement.").

17. BISHOP & KLEINBERGER, *supra* note 8, at ¶ 13.01[1] (emphasis added).

18. ULLCA § 103(b) (1996).

19. *Id.* § 103(a).

20. *See* BISHOP & KLEINBERGER, *supra* note 8, at ¶ 13.01[1] n.5 (Hawaii, South Carolina, Vermont, and West Virginia).

21. Andree Sophia Blumstein, *Key Tax Aspects of the Tennessee Limited Liability Company Act*, TENN. B.J., Jul./Aug. 1994, at 14, 14; *see also* Tennessee Limited Liability Company Act, Tenn. Code Ann. §§ 48-201-101 to 48-248-606 (2002).

22. *See* TENNBARU, *supra* note 2, at 1; *see also supra* Part II.A.

"opted for an LLC act with technical requirements that would force lawyers to organize LLCs that would be treated as partnerships for federal income tax purposes."[23] Since the Prior Act passed in 1994, it has been subject to several amendments. While the Prior Act has served its purpose well, subsequent amendments which have attempted to keep the Prior Act current with federal tax rules, have created somewhat of a monster.[24]

Like LLC statutes in many other states,[25] the Prior Act was based on a conglomeration of concepts taken from corporation and partnership law.[26] In fact, "[m]any of the [Prior Act's] provisions were taken virtually verbatim from the analogous sections of the Corporate Act, the Partnership Act, or the Limited Partnership Act."[27] Like the New Act, the Prior Act contains many default provisions that are simply default provisions.[28] Additionally, the Prior Act contains several nonwaivable provisions that are scattered throughout various chapters and parts of the Prior Act.[29] Throughout the Prior Act, default provisions are signaled by language such as "[u]nless otherwise provided in the articles or operating agreement."[30] Nonwaivable provisions are indicated by the use of a mandatory verb such as *must* or *shall*.[31] The Prior Act, however, provides little overall guidance for sorting through and distinguishing the default and nonwaivable provisions.[32]

The Prior Act allows for two possible LLC management paradigms: (1) *member*-managed and (2) *board*-managed.[33] The Prior Act requires the LLC articles of organization to "set forth . . . [a] statement as to whether the LLC will be board-managed or whether the LLC will be member-managed."[34] A Tennessee member-managed LLC under the Prior Act possesses many of the

---

23. TENNBARU, *supra* note 2, at 1.

24. *Id.* The Prior Act has a total of 181 sections while the Delaware LLC Act, has 83 sections and the ULLCA has 86 sections. *Id.* Tennessee's New Act contains only 136 sections. *See* Tenn. Code Ann. §§ 48-249-101 to -249-1133 (Supp. 2005).

25. *See, e.g.*, J. WILLIAM CALLISON & MAUREEN A. SULLIVAN, LIMITED LIABILITY COMPANIES: A STATE-BY-STATE GUIDE TO LAW AND PRACTICE § 15:13 (2005) (discussing the origin of Arizona LLC law); Carney, *supra* note 4, at 858 (discussing the origin of Wyoming LLC law).

26. RICHARD R. SPORE, III, BUSINESS ORGANIZATIONS IN TENNESSEE 5-1 (2d ed. 2003).

27. *Id.* at 5-2.

28. *See, e.g.*, TENN. CODE ANN. § 48-239-107(a) (2002).

29. *See, e.g., id.* § 48-205-101(1)-(16) (stating what the articles of incorporation must include).

30. *See, e.g.*, § 48-239-107(a).

31. *See, e.g.*, § 48-205-101 (stating that "[t]he articles *must* set forth" certain specific information (emphasis added)).

32. Tennessee legal scholars have made the task of sorting through these provisions easier by creating forms and treatises that centralize key nonwaivable provisions. *See, e.g.*, 14A LESLIE B. WILKINSON, JR., TENNESSEE PRACTICE 13:15 (Nancy Fraas MacLean ed., 2002).

33. TENN. CODE ANN. § 48-238-101(a)(2) (2002) ("An LLC shall be either member-managed or board-managed . . . .").

34. *Id.* § 48-205-101(5).

same governance traits of a Delaware member-managed LLC.[35] Departing from the traditional approach, however, a board of governors manages a board-managed LLC.[36]

The Prior Act also requires every Tennessee LLC to appoint a chief manager and a secretary to serve as officers of the company.[37] This requirement is mandatory and nonwaivable in both the member-managed and board-managed forms.[38] If a chief manager or secretary are not appointed, "the person or persons exercising the principal functions of the chief manager or the secretary are considered to have been elected to those offices."[39] Other managers and officers may be appointed or elected in either management form.[40]

The chief manager has the same duties in either the member-managed or board-managed LLC.[41] The Prior Act states that the chief manager shall perform the following duties:

> (1) See that all orders and resolutions of the board of governors or members are carried into effect;
> (2) Sign and deliver in the name of the LLC any deeds, mortgages, bonds, contracts or other instruments pertaining to the business of the LLC, except in cases in which the authority to sign and deliver is required by law to be exercised by another person or is expressly delegated:
> (A) By the articles or operating agreement;
> (B) The board of governors if the LLC is board-managed; or
> (C) Members if the LLC is member-managed to some other manager or agent of the LLC;
> (3) Perform other duties prescribed by the board of governors or the members; and
> (4) In the event the LLC has a vacancy in the office of secretary, any notices, documents or other matters that otherwise are required to go to the secretary

---

35. *Compare* TENN. CODE ANN. § 48-238-101(a)(1) (2002) ("[A]ll powers shall be exercised by or under the authority of, and the business and affairs of the LLC shall be managed by or under the direction of its members."), *with* DEL. CODE ANN. tit. 6, § 18-402 (1999) ("[T]he management of a limited liability company shall be vested in its members . . . .").

36. *Compare* TENN. CODE ANN. § 48-238-101(a)(2) (2002) ("An LLC shall be either member-managed or board-managed . . . ."), *with* DEL. CODE ANN. tit. 6, § 18-402 (1999) ("[I]f a limited liability company agreement provides for the management . . . by a manager, the management . . . shall be vested in the manager . . . ."). The Prior Act requires that the board of governors be comprised of natural persons. TENN. CODE ANN. § 48-239-102(a) (2002). While this requirement appears to be a nonwaivable provision due to the use of "must" in the statute, an entity that is a member of the board-managed LLC may also be a governor if a natural person, who is an officer or agent of that entity, "function[s] on behalf of the entity." *Id.* § 48-239-102(c).

37. TENN. CODE ANN. § 48-241-101 (2002).

38. *See id.*

39. *Id.* § 48-241-106.

40. *Id.* § 48-241-103.

41. *See id.* § 48-241-102(b). While the duties are practically the same, the authority to carry out those duties is derived from either the board or members depending on the form of management. *See id.*

may be delivered to the chief manager.[42]

As demonstrated above, the chief manager is, in fact, the chief *agent* of both member-managed and board-managed LLCs. One of the key distinctions, however, between member-managed and board-managed LLCs is that, by default, members retain agency authority in a member-managed LLC[43] but have no agency authority in a board-managed LLC.[44] The agency authority of a member in a member-managed LLC is limited to actions "for apparently carrying on in the usual way [of] the business of the LLC."[45] Note, however, that this is only a default rule that can be modified by the articles of incorporation.[46] Furthermore, even if the agency status of a member is so modified, the member retains apparent authority to act as an agent for the LLC.[47]

In contrast, governors in a board-managed LLC under the Prior Act have no authority as agents.[48] As with the agency rules of a member-managed LLC, this is a default provision that may be modified by the articles of incorporation, the operating agreement, or by action of the governors.[49] If no other agents are designated, the chief manager is the sole agent of a board-managed LLC.[50] Furthermore, *members* of a board-managed LLC, unlike the agency rules for a member-managed LLC, have no default actual or apparent authority to bind the LLC even if "carrying on in the usual way the business of the LLC."[51]

While nuances between member-managed and board-managed LLCs can seem confusing, these two management paradigms under the Prior Act can be equated with a partnership and a corporation, respectively. A Tennessee member-managed LLC under the Prior Act, like those in other states, functions very much like a partnership with the exception of requiring a chief manager. A Tennessee board-managed LLC under the Prior Act, unlike the manager-managed LLC which is the common complementary paradigm to member-managed LLCs, functions much like a corporation. The New Act, however, eliminates board-managed LLCs, but provides for a similar management paradigm (the "director-managed" LLC)[52] for those who have grown accustomed to or see the benefits of Tennessee's board-managed structure. Furthermore, the New Act unifies Tennessee LLC law with many of the popular aspects of LLC law throughout the

---

42. *Id.*

43. *Id.* § 48-238-103(a).

44. *Id.* § 48-238-104(a).

45. *Id.* § 48-238-103(a).

46. *Id.*

47. *Id.* ("[T]he act of every member . . . for apparently carrying on in the usual way the business of the LLC . . . binds the LLC, unless the member so acting has in fact no authority to act for the LLC in the particular matter, and the person with whom the member is dealing has knowledge of the fact that the member has no such authority.").

48. *Id.* § 48-238-104(b).

49. *Id.*

50. *See id.*

51. *Id.* § 48-238-104(a).

52. *See* discussion *infra* Part III.A.

rest of the nation.

C. An Introduction to the Tennessee Revised Limited Liability Company Act

The New Act applies to all LLCs formed after January 1, 2006.[53] LLCs formed prior to January 1, 2006 will continue to be governed by the Prior Act, unless they expressly choose to be governed by the New Act.[54] There is no sunset provision regarding the applicability of the Prior Act to LLCs formed before January 1, 2006.[55] As a result, both the New Act and Prior Act will govern LLCs in Tennessee depending on the particular LLC at issue.[56]

The New Act was drafted with a "strong bias toward freedom of contract."[57] As noted above, freedom of contract is a key feature of LLCs in the United States.[58] This policy of freedom of contract is evidenced in (among other things) a provision permitting an oral operating agreement,[59] the reduction of nonwaivable provisions,[60] and the ability of any other entity to convert to an LLC.[61]

Naturally, an oral operating agreement brings with it potential pitfalls for the unwary drafter. For example, an operating agreement may be drafted to regulate certain affairs of the LLC.[62] However, because "[t]he written provisions of an operating agreement need not be set out in a single integrated document,"[63] a subsequent oral operating agreement could potentially modify the written operating agreement. Attorneys should recognize this potential pitfall and avoid this hazard by drafting articles of incorporation that expressly prohibit the use of oral operating agreements.[64]

---

53. TENN. CODE ANN. § 48-249-1002(a)(1)(A) (Supp. 2005).

54. *Id.* § 48-249-1002(b).

55. *Id.* § 48-249-1002(c). This is not an uncommon occurrence in Tennessee business organization law. When the Tennessee Revised Limited Partnership Act took effect in 1989, limited partnerships formed prior to January 1, 1988 were not subject to the new legislation unless the limited partnership filed "a certificate of amendment which would cause its certificate of limited partnership to comply with [the Revised Limited Partnership Act] and which specifically states that it is electing to be so bound and by paying the fee for a certificate of limited partnership." TENN. CODE ANN. § 61-2-1204(c), (e) (2002).

56. *See* TENN. CODE ANN. § 48-249-1002(a)-(c) (Supp. 2005).

57. Samuel E. Stumpf, Understanding the New Limited Liability Company Act, Address at TennBarU seminar (June 29, 2005).

58. *See supra* text accompanying note 9.

59. TENN. CODE ANN. § 48-249-203(a) (Supp. 2005).

60. *Compare* TENN. CODE ANN. §§ 48-201-101 to 48-248-606 (2002) (incorporating more than twenty-five nonwaivable provisions), *with* TENN. CODE ANN. § 48-249-205(b) (Supp. 2005) (incorporating only twenty-one nonwaivable provisions).

61. TENN. CODE ANN. § 48-249-703(a), (h) (Supp. 2005).

62. *Id.* § 48-249-203(a).

63. *Id.*

64. *See id.* (suggesting that a requirement in the articles of organization or a *written* provision

One of the benefits of the oral operating agreement permits Tennessee's "mom and pop" establishments to agree on how "to regulate the affairs of the LLC and the conduct of its business"[65] without the expense of legal advice. While permitting an oral operating agreement may seem like a drastic change, one should note that Tennessee partnership law has always permitted an oral agreement to govern the affairs of a partnership.[66] Furthermore, both Delaware LLC law and the ULLCA permit the use of oral operating agreements.[67] As long as attorneys recognize this issue, business clients will retain the control they seek while do-it-yourselfers can now efficiently operate their LLC.

To assist with and promote freedom of contract, the New Act consolidates all of the nonwaivable provisions into a single section.[68] Therefore, any provision of the New Act that does not appear in this consolidated list is waivable.[69] The section of the New Act consolidating these nonwaivable provisions is consistent with the ULLCA.[70] In fact many provisions of the New Act are drawn largely from the ULLCA or the Delaware LLC Act,[71] which are the two most influential acts for the Tennessee LLC revisions.[72]

Another substantial change in the New Act allows any type of entity to convert to a Tennessee LLC.[73] Under the Prior Act, only a partnership or limited partnership could convert to an LLC.[74] Additionally, the New Act permits a broader scope of entities to convert to LLCs than either Delaware[75] or the ULLCA.[76] Note, however, that one limitation on conversion under the New Act

---

in an operating agreement may mandate that all operating agreements be in writing).

65. *Id.*

66. *See* TENN. CODE ANN. § 61-1-103(b) (2002) ("[R]elations among the partners and between the partners and the partnership are governed by the partnership agreement"); *Id.* § 61-1-101(7) ("'Partnership agreement' means the agreement, whether written, *oral*, or implied, among the partners concerning the partnership, including amendments to the partnership agreement." (emphasis added)).

67. *See* DEL. CODE ANN. tit. 6, § 18-101(7) (1999); ULLCA § 103 (1996).

68. *See* TENN. CODE ANN. § 48-249-205(b) (Supp. 2005).

69. *Id.* § 48-249-205(a). This provision of the New Act creates some interesting features for LLCs. For example, a director *must* be an individual. *Id.* § 48-249-205(b)(1). A corporation (or any other entity), however, may be an officer of the LLC if provided for in the operating agreement or articles of organization. *See id.* § 48-249-102(22) (requiring an officer to be an individual); *Id.* § 48-249-205(b) (failing to identify § 102(22) as a nonwaivable provision).

70. *Compare* TENN. CODE ANN. § 48-249-205 (Supp. 2005), *with* ULLCA § 103 (1996).

71. Limited Liability Company Act, DEL. CODE ANN. tit. 6, §§ 18-101 to -1109 (1999).

72. *See* TENNBARU, *supra* note 2, sched. II at 31 ("Rules of Engagement" for the Committee).

73. *See* TENN. CODE ANN. § 48-249-703 (Supp. 2005) (setting forth the terms of how "[a]ny other entity may convert to a [Tennessee] LLC").

74. *See* TENN. CODE ANN. § 48-204-101(a) (2002) ("A general or limited partnership organized in this state may be converted to an LLC pursuant to this section.").

75. *See* DEL. CODE ANN. tit. 6, § 18-214 (1999) (permitting specific enumerated entities to convert).

76. *See* ULLCA § 902(a) (1996) (permitting only partnerships and limited partnerships to

is that the applicable law governing the entity seeking conversion must permit that entity to convert.[77]

Another new and prominent feature of the New Act is the ability to establish "series of members, holders, managers, directors, membership interests or financial rights [that have] separate rights, powers and/or duties" from the LLC.[78] The practical effect of this "series" provision allows for an LLC arrangement similar to the parent and subsidiary relationship in corporation law.[79] While an in-depth discussion of this new feature is beyond the scope of this Comment, it is worthwhile to note that this "series" section in the New Act was drawn from the Delaware LLC Act.[80] Although the language in the New Act is much more condensed than the comparable section of the Delaware LLC Act, the New Act's "series" provision is intended to have the same substantive effect as Delaware's "series" statute.[81]

Delaware courts have not addressed the effects of the "series" provision, and consequently, concern has been raised regarding the effect an LLC series has on creditors' rights.[82] One such concern is whether a "separate series [will] be respected by other states for liability purposes?"[83] In light of these concerns, care should be exercised in applying the New Act's "series" provision. Attorneys should monitor the Delaware courts' position on their state's "series" statute to gain some insight on how the Tennessee courts might interpret the New Act.

While reliance on the ULLCA and the Delaware LLC Act is evident throughout the New Act, many provisions that are special and unique to

---

convert).

77. *See* TENN. CODE ANN. § 48-249-703(g) (Supp. 2005).

78. *See id.* § 48-249-309(a).

79. Julia Gold, *Series Limited Liability Companies - Too Good to be True?*, NEV. LAW., July 2004, at 18.

80. TENNBARU, *supra* note 2, at 7; *see* DEL. CODE ANN. tit. 6, § 18-215 (1999).

81. *See* TENNBARU, *supra* note 2, at 7. One treatise discussing the Delaware "series" statute noted the following:

> The series arrangement is most advantageous where an LLC has several valuable assets, each of which has associated with it different magnitudes of actual or potential liabilities. In LLCs other than Delaware LLCs, a creditor cannot enforce an obligation of the LLC against the members of the LLC but that creditor can enforce the obligation against all of the assets of the LLC, including those assets that were not the subject of the particular obligation. The only means by which such an LLC can protect valuable assets from creditors while maintaining overall control of the assets is to create separate LLCs that are linked through mutual ownership of membership interests. Creating and maintaining separate LLCs can cause expensive and inefficient administrative burdens. The series arrangement eliminates the need to create and maintain separate LLCs to protect discrete assets.

NICHOLAS KARAMBELAS, LIMITED LIABILITY COMPANIES: LAW, PRACTICE AND FORMS § 7.4 (2005).

82. Gold, *supra* note 79, at 20.

83. *Id.* (suggesting that "it would be good business practice to provide all creditors actual notice of the existence of the separate series thereby eliminating any notice argument a creditor may have").

Tennessee remain in the New Act or were modified with state-specific concerns in mind.[84] This combination resulted in an Act that while featuring proven and agreed-upon LLC provisions, is tailored to the specific needs and concerns of Tennessee businesses and attorneys.

III. THE THREE MANAGEMENT PARADIGMS OF THE NEW LIMITED LIABILITY COMPANY ACT

One of the most intriguing and unique features of the New Act is the capability to organize an LLC as one of three different management paradigms[85]—a choice available only under Tennessee law. Under the director-managed paradigm, "LLC powers [are] exercised under the authority of . . . [a] board of directors," much like a corporation.[86] Under the manager-managed paradigm, the traditional LLC form is revisited; "[e]ach manager has equal rights in the management and conduct of the LLC's business," much like a limited partnership.[87] Finally, under the member-managed paradigm, "[e]ach member has equal rights in the management and conduct of the LLC's business," much like a partnership.[88]

A. The Director-Managed LLC

"In a director-managed LLC . . . [a]ll LLC powers shall be exercised under the authority of, and the business and affairs of the LLC shall be managed under the direction of, its board of directors . . . ."[89] The structure of this new management paradigm—a true corporate management scheme - replaces the board-managed paradigm under the Prior Act. The New Act requires a director-managed LLC to "have a president who is appointed or elected by a majority vote of the directors and authorized to act as an agent of the LLC."[90] This president is the only statutorily required officer for a director-managed LLC.[91]

The nomenclature used for parties involved in director-managed LLCs is consistent with other areas of business organization law and other jurisdictions. The "governors" and "chief manager" titles under the Prior Act have been replaced with "directors" and "president" titles, respectively, under the New Act.[92] Additional officers are simply that—officers.[93] In fact, the "manager" title

84. TENNBARU, *supra* note 2, sched. II at 31 ("Rules of Engagement" for the Committee).

85. TENN. CODE ANN. § 48-249-401(a)-(c) (Supp. 2005).

86. *Id.* § 48-249-401(c) (Supp. 2005).

87. *Id.* § 48-249-401(b).

88. *Id.* § 48-249-401(a).

89. *Id.* § 48-249-401(c)(1).

90. TENN. CODE ANN. § 48-249-401(d) (Supp. 2005).

91. *See id.* § 48-249-401(c)-(d).

92. *Compare* Tenn. Code Ann. §§ 48-238-101(a)(2), 48-239-102(c) (2002) (requiring that governors be appointed in a board-managed LLC and that a chief manager be appointed in every LLC), *with* TENN. CODE ANN. § 48-249-401(c)-(d) (Supp. 2005) (requiring the appointment of

is excluded entirely from the director-managed LLC.[94]

In a director-managed LLC, "no member or director is an agent of the LLC for the purpose of its business solely by reason of being a member or a director."[95] Only "[t]he president and any other authorized officers of a director-managed LLC shall be agents of the LLC."[96] The president has actual and apparent authority "for carrying on in the ordinary course the LLC's business or business of the kind carried on by the LLC."[97] Additionally, "knowledge" and "notice" for apparent authority purposes are defined in the New Act.[98]

Members of a director-managed LLC "owe[] no duties to the LLC or to the other members or holders of financial rights solely by reason of being a member."[99] The New Act simply requires a director to "discharge all duties as a director . . . [i]n good faith . . . [w]ith the care an ordinarily prudent person in a like position would exercise under similar circumstances; and . . . [i]n a manner the director reasonably believes to be in the best interests of the LLC."[100] The president and any other appointed officers are held to essentially the same standards and duties as a director.[101]

B. The Manager-Managed LLC

The manager-managed LLC is entirely new to Tennessee business organization law. What has long been a traditional management paradigm in other jurisdictions is now available to those in Tennessee who wish to create an LLC with a governance structure comparable to a limited partnership. Following the traditional structure of a manager-managed LLC, "[e]ach manager [in a manager-managed LLC] has equal rights in the management and conduct of the LLC's business."[102] A manager does not have to be a member of the LLC.[103] With some exceptions,[104] "any matter relating to the business of the LLC shall be exclusively decided by the manager or, if there is more than one (1) manager, by a majority vote of the managers."[105]

As with the director-managed LLC, the titles used in the manager-managed LLC are traditional titles, and those familiar with manager-managed LLCs in

---

director(s) and a president in a director-managed LLC).

93. *See* TENN. CODE ANN. § 48-249-102(23) (Supp. 2005).

94. *See id.* § 48-249-401(c)-(d).

95. *Id.* § 48-249-402(c).

96. *Id.*

97. *Id.* § 48-249-402(d).

98. *Id.* § 48-249-402(f).

99. *Id.* § 48-249-403(i)(1).

100. *Id.* § 48-249-403(i)(4).

101. *See id.* § 48-249-403(j).

102. *Id.* § 48-249-401(b)(1).

103. *Id.* § 48-249-401(b)(3)(C).

104. *Id.* § 48-249-401(e) (delegation of management rights and powers); *Id.* § 48-249-401(f) (unanimous consent required for certain matters).

105. *Id.* § 48-249-401(b)(2).

other jurisdictions will readily recognize their form and function. Members and managers are the two key titles and positions in a manager-managed LLC, which function as limited and general partners in a limited partnership, respectively.[106] Additionally, a president and other officers or agents may be appointed, much like the LLCs under the Prior Act.[107]

Managers in the manager-managed LLC are *the* agents of the LLC.[108] Furthermore, "[a] member is not an agent of the LLC for the purpose of its business solely by reason of being a member."[109] A manager retains actual and apparent authority "for carrying on in the ordinary course the LLC's business or business of the kind carried on by the LLC."[110]

The New Act holds managers "to the same standards of conduct prescribed for members" in member-managed LLCs.[111] Additionally, "[a] member owes no duties to the LLC or to the other members or holders of financial rights solely by reason of being a member."[112] Similar to a limited partner exercising the powers of a general partner,[113] "[a] member [in a manager-managed LLC] that . . . exercises some . . . of the rights of a manager in the management and conduct of the LLC's business is held to the standards of conduct" of a manager.[114]

C. *The Member-Managed LLC*

The member-managed LLC is the only management paradigm found in both the Prior Act and the New Act.[115] The form and function of this management paradigm remain practically the same.[116] Accordingly, the New Act states that "[i]n a member-managed LLC . . . [e]ach member has equal rights in the management and conduct of the LLC's business."[117] Additionally, "[e]ach member is an agent of the LLC for the purpose of its business."[118]

Under the New Act, "[t]he *only* fiduciary duties a member owes to a member-managed LLC and the LLC's other members and holders are the duty of loyalty and the duty of care."[119] The duty of loyalty is limited to three factors: (1)

106. *Id.* § 48-249-401(b).
107. *Id.* § 48-249-401(e).
108. *Id.* § 48-249-401(b)(2); *id.* 48-249-402(b)(1).
109. *Id.* § 48-249-402(b)(1).
110. *Id.*
111. *Id.* § 48-249-403(h)(2).
112. *Id.* § 48-249-403(h)(1).
113. TENN. CODE ANN. § 61-2-302(a) (2002).
114. TENN. CODE ANN. § 48-249-403(h)(3) (Supp. 2005).
115. *See supra* text accompanying notes 32-35.
116. Notably absent from this paradigm, however, is a provision requiring the appointment of a chief manager. *Compare* TENN. CODE ANN. § 48-241-101 (2002) (requiring the appointment of a chief manager), *with* TENN. CODE ANN. § 48-249-401(a) (Supp. 2005) (describing a member-managed LLC without mentioning a chief manager). *See supra* Part II.B.
117. TENN. CODE ANN. § 48-249-401(a)(1) (Supp. 2005).
118. *Id.* § 48-249-402(a)(1).
119. *Id.* § 48-249-403(a) (emphasis added).

properly accounting for and using the LLC's property, profits, and opportunities;[120] (2) abstaining from dealings with those who have an adverse interest to the LLC;[121] and (3) "refrain[ing] from competing with the LLC."[122] The duty of care in a member-managed LLC "is limited to refraining from engaging in grossly negligent or reckless conduct, intentional misconduct or a knowing violation of law."[123]

The statutory limitation on fiduciary duties is an important protection from courts' attempts to apply common law fiduciary duties akin to those already existing in Tennessee corporation law.[124] Prior to the adoption of the New Act, the Tennessee Court of Appeals in *Anderson v. Wilder*[125] applied corporate common law and concluded that members owning a majority interest in a member-managed LLC owe a fiduciary duty to members owning a minority interest.[126] The court noted that, because LLCs are hybrids of corporations and partnerships, LLCs should carry the same characteristics of these entities "with emphasis on the foundational business form from which that characteristic originated."[127] The court cited to the statutory fiduciary duties of partners in a partnership,[128] which are nearly verbatim to the New Act's fiduciary duties for members and limits a partner's fiduciary duties to those enumerated in the statute.[129] Additionally, the court noted the fiduciary duties owed by majority shareholders to minority shareholders in closely-held corporations, which are entirely common law.[130] The court found that it is essential for this same fiduciary duty to exist among the members of an LLC because it is a universal duty in both partnerships and closely-held corporations.[131]

The Tennessee Court of Appeals noted that its decision to apply corporate common law fiduciary duties to LLCs did not conflict with the Prior Act's fiduciary duty statute.[132] By doing so, the court implied that the absence of language in the Prior Act limiting the fiduciary duties to those specifically

120. *Id.* § 48-249-403(b)(1).

121. *Id.* § 48-249-403(b)(2); *see also id.* § 48-249-404 (defining "conflict of interest transaction" and discussing the scope of the duty of loyalty).

122. *Id.* § 48-249-403(b)(3).

123. *Id.* § 48-249-403(c). In contrast, a member in a manager-managed LLC under the New Act "owes *no duties* to the LLC or to the other members or holders of financial rights solely by reason of being a member." *Id.* § 48-249-403(h)(1) (emphasis added).

124. *See* Richard R. Spore, III, Understanding the New Limited Liability Company Act, Address at TennBarU seminar (June 29, 2005).

125. No. E2003-00460-COA-R3-CV, 2003 WL 22768666 (Tenn. Ct. App. Nov. 21, 2003).

126. *Id.* at *6.

127. *Id.* at *4.

128. *Id.* at *4-5.

129. TENN. CODE ANN. § 48-249-403(b) (Supp. 2005); *cf.* TENN. CODE ANN. § 61-1-404 (2002) (setting forth the same duties for partners in a partnership that the New Act applies to members in a member-managed LLC).

130. *See Anderson* at *3.

131. *See id.*

132. *See id.*

enumerated in the statute, like a partner's duty in a partnership, permitted the judiciary's application of common law closely-held corporation fiduciary duties to members of an LLC.[133] With the Tennessee Court of Appeals's reasoning in mind, the drafters of the New Act responded to *Anderson*[134] by spelling out the precise fiduciary duties for members and other parties who are part of Tennessee LLCs.[135]

IV. CONCLUSION

The recent passage of the New Act is a substantial and necessary change in Tennessee business organization law. The streamlined structure and uniform approach of the New Act should prove helpful for businesses and attorneys alike. Furthermore, organizing an LLC in Tennessee should be more accommodating to both "mom and pop" establishments, who want to keep it simple, and larger businesses, who organize more complex LLCs with the guidance of counsel.[136]

Tennessee now has an LLC Act that is more unified with the rest of the country. Drawing heavily from the ULLCA, Delaware law, and what the drafters believed were necessary components of existing Tennessee law,[137] the New Act certainly appears to be reliable. Additionally, the New Act's uniformity with other states' laws may assist Tennessee courts that are searching other jurisdictions for analogous LLC controversies. Finally, the New Act advances the reform in LLC law sweeping the nation by providing three rather than two management paradigms[138]—an advancement that other jurisdictions may adopt in the future.

ERIC REAGAN

133. *See* TENN. CODE ANN. § 48-240-102 (2002); *cf.* TENN. CODE ANN. § 61-1-404 (limiting a partner's fiduciary duties to *only* those duties specifically enumerated in the statute).

134. *See* Spore, *supra* note 124.

135. TENN. CODE ANN. § 48-249-403(a)-(c) (Supp. 2005).

136. *See supra* text accompanying notes 63-65.

137. *See supra* Part II.C.

138. *See supra* Part III (discussing the three management LLC paradigms available in Tennessee under the New Act).

CONSTITUTIONAL LAW—STANDARD OF REVIEW FOR FOURTEENTH AMENDMENT CHALLENGES–STRICT SCRUTINY IS THE STANDARD OF REVIEW FOR EQUAL PROTECTION CHALLENGES TO GOVERNMENTAL RACIAL CLASSIFICATION IN PRISON

Johnson v. California, 543 U.S. 499 (2005).

I. INTRODUCTION

On June 22, 1987, plaintiff, an African-American male, began serving a sentence of twenty-five years to life in the California Institution for Men in Chino, California, for the crimes of assault with a deadly weapon, robbery, and murder.[1] During plaintiff's initial evaluative period, the California Department of Corrections ("CDC") housed him with an African-American roommate in an inmate reception center.[2] An unwritten CDC reception center policy prevented new or transferred inmates, under almost all circumstances, from being assigned a cellmate of a different race during the evaluative period.[3] Because the policy applied to transferred inmates as well as new inmates, the CDC had racially segregated plaintiff five times.[4]

---

1. Johnson v. California, 543 U.S. 499, 525 (2005) (Thomas, J., dissenting).

2. *Id.* at 503 (majority opinion); Brief for Petitioner at 2 n.1, Johnson v. California, No. 03-636 (June 3, 2004). The CDC evaluates prisoners during the "reception" period, which lasts up to sixty days. *Johnson*, 543 U.S. at 525 (Thomas, J., dissenting). During the reception period, prison officials determine an inmate's security needs and permanent cell assignment based upon health, criminal history, presence of enemies in prison, physical, mental, and emotional health. *Id.* at 525-26. Frequently, a prisoner's pre-sentence reports, which may indicate gang affiliation and criminal history, arrive several weeks after the inmate arrives. *Id.* at 522 (Stevens, J., dissenting). Therefore, the testing process becomes increasingly important for assessing the compatibility of an inmate with potential permanent cellmates. *See id.* at 525 (Thomas, J., dissenting).

3. *Id.* at 502 (majority opinion). Race is the predominant factor in double-cell assignments in reception centers. *Id.* The CDC usually subdivides prisoners within each racial group. *Id.* One method is to subdivide based upon inmates' regional origin within the United States. *Id.* Therefore, Southern California Hispanics may be separated from the Northern California Hispanics. *Id.* Alternatively, inmates may be subdivided by the national origin of the inmates' ethnic group. *Id.* For example, Chinese-Americans may be separated from the Japanese-Americans. *Id.* The CDC asserts that this temporary racial segregation prevents interracial violence between members of different gangs. *Id.*

4. Brief for Petitioner, *supra* note 2, at 2. "In 2003, the [California reception] centers processed more than 40,000 newly admitted inmates, almost 72,000 inmates return[ing] from parole, over 14,000 inmates admitted for other reasons, and some portion of the 254,000

On February 24, 1995, plaintiff filed a complaint *pro se* in the Federal District Court for the Central District of California.[5] He alleged the defendants'[6] policy of assigning cellmates in reception centers on the basis of race violated his constitutional right to equal protection under the Fourteenth Amendment.[7] The district court granted summary judgment in favor of the defendants, on the grounds that they were entitled to qualified immunity because their actions were "not clearly unconstitutional."[8] On February 25, 2003, the appellate court rejected the district court's qualified immunity analysis and turned to the merits of the case to determine whether a constitutional right had been violated.[9] The Ninth Circuit Court reviewed the

---

inmates who were transferred from one prison to another." *Johnson*, 543 U.S. at 525 (Thomas, J., dissenting).

5. Johnson v. California, 207 F.3d 650, 653 (9th Cir. 2000) (describing the case's prior history in district court). The United States District Court for the Central District of California provides copies of decisions and most orders for five years on their home page. United States District Court for the Central District of California, *available at* http://www.cacd.uscourts.gov/. Plaintiff's district court hearings are not published. *See id.*

6. *Johnson*, 207 F.3d at 653. Plaintiff named as defendants the former CDC Director James Rowland and James Gomez, CDC Director from 1991 until the time plaintiff filed his complaint, alleging they administered an unconstitutional policy of assigning cellmates on the basis of race during their respective tenures. *See id.*

7. *Johnson*, 543 U.S. at 503. *See* U.S. CONST. amend. XIV, § 1. Plaintiff alleged that the defendants housed inmates on the basis of race knowing that the practice incited racial tension because they were paid at a higher rate during racial riots. *Johnson*, 207 F.3d at 655. Additionally, plaintiff alleged that defendants conspired to extort money from the inmates by overcharging for telephone calls. *Id.* at 656. He asserted the defendants violated his right to due process under the Fifth Amendment [and] his right to be free from cruel and unusual punishment under the Eighth Amendment. *Id.* at 653.

The district court magistrate judge granted plaintiff leave to amend his *pro se* complaint three times. *Id.* at 654. In January of 1998, the district court dismissed plaintiff's complaint for failure to state a claim. *Id.* at 653. On March 21, 2000, the Ninth Circuit Court of Appeals reversed and remanded the district court's decision, holding that plaintiff's claims were "sufficient to state a claim for racial discrimination in violation of the Equal Protection Clause of the Fourteenth Amendment." *Id.* at 653, 655. The appellate court stated that "the rule of liberal construction [of complaints to determine the merit of allegations] is particularly important in civil rights cases." *Id.* at 653 (quoting Ferdik v. Bonzelet, 963 F.2d 1258, 1261 (9th Cir. 1992)).

8. Johnson v. California, 321 F.3d 791, 795 (9th Cir. 2003). The district court applied the United States Supreme Court's analysis for qualified immunity under Saucier v. Katz, 533 U.S. 194 (2001). Under the *Saucier* analysis, the initial inquiry is whether the facts alleged show that conduct violated a constitutional right. *Id.* at 201. The second inquiry is whether the right violated was "clearly established" in the context of the case. *Id.* The district court found that under this two part test, the defendants "were entitled to qualified immunity because their actions were not clearly unconstitutional." *Johnson*, 321 F.3d at 795.

9. *Johnson*, 321 F.3d at 807. The Ninth Circuit found that before addressing the issue of qualified immunity, the court must first determine that the defendants' conduct actually violates another person's constitutional rights under the *Saucier* analysis. *Id.* In *Johnson*, the Ninth Circuit found no conduct by the defendants violated constitutional rights; therefore, there was no need to examine whether defendants were entitled to qualified immunity. *Id.*

policy of race-based cell assignments under the deferential standard set forth in *Turner v. Safley*.[10] The defendants asserted that the prison policy was implemented in the interest of prison security.[11] The court found that plaintiff did not carry his burden of proving that the temporary racial segregation policy was not reasonably related to the penological interest of preventing racial violence.[12] Thus, in the absence of evidence that the policy was unreasonable, the Ninth Circuit deferred to the defendants' expertise.[13]

On certiorari to the United States Supreme Court, *held*, reversed and remanded.[14] Strict scrutiny is the proper standard of review for an equal protection challenge to a prison policy using racial classification. *Johnson v. California*, 543 U.S. 499 (2005).

The issue before the Court in *Johnson* was whether strict scrutiny or deferential review is the standard of review for defendants' race-based prison policy.[15] Under *Turner,* courts use deferential review when determining whether a prison policy that burdens an inmate's constitutional rights reasonably relates to legitimate penological interests.[16] In *Adarand Constructors, Inc. v. Pena*, however, the Court held that government-imposed racial classifications must be analyzed under a strict scrutiny analysis.[17] Defendants argued the United States Supreme Court should review the CDC policy under *Turner*'s deferential standard because wardens have specialized knowledge related to managing prison populations.[18] Plaintiff argued that the deferential standard should not extend to equal protection challenges of race-based policies in prison.[19] Strict scrutiny applies to governmental policies that impinge on the

---

10. *Id.* at 798-99; Turner v. Safley, 482 U.S. 78, 89-90 (1987).

11. *Johnson*, 321 F.3d at 799.

12. *See id.* at 799, 807 ("Under *Turner,* rather than the administrators bearing the burden of proving their housing policy constitutional, the inmate bears this 'heavy burden': To prevail, [Johnson] must overcome the presumption that the prison officials acted within their broad discretion.") (alteration in original) (quoting Shaw v. Murphy, 532 U.S. 223, 232 (2001)).

13. *See id.*; *see also* Julie Taylor, *Racial Segregation in California Prisons*, 37 LOY. L.A. L. REV. 139, 152 (2003) (discussing the Ninth Circuit's analysis in *Johnson*). On July 28, 2003, the Ninth Circuit denied plaintiff's petition for rehearing en banc. *Johnson*, 321 F.3d 791 (9th Cir. 2003) *reh'g denied per curiam*, 336 F.3d 1117 (9th Cir. 2003). Judge Ferguson dissented, outlining two principle objections to a deferential standard of review in the context of racial classification in prison. *Johnson*, 336 F.3d at 1117-23 (Ferguson, J., dissenting). First, Judge Ferguson asserted that "[t]he panel's decision ignores the Supreme Court's repeated and unequivocal command that all racial classifications imposed by the government must be analyzed by a reviewing court under strict scrutiny." *Id.* at 1117 (internal quotation marks omitted) (quoting Adarand Constructors, Inc. v. Pena, 515 U.S. 200, 227 (1995)). Second, Judge Ferguson argued that *Turner* should not apply in Johnson's case because "the [Fourteenth Amendment] right asserted is not inconsistent with legitimate penological objectives." *Id.*

14. Johnson v. California, 543 U.S. 499, 501 (2005).

15. *Id.* at 504-05.

16. *See* Turner v. Safley, 482 U.S. 78, 89-90 (1987).

17. 515 U.S. 200, 201 (1995).

18. *Johnson*, 543 U.S. at 509-12.

19. *See id.* at 503-05.

right to equal protection because the standard "smoke[s] out invidious racial classifications,[20] and Plaintiff argued that the CDC policy should not be exempt from this "categorical strict scrutiny rule."[21] In *Johnson*, the United States Supreme Court decided whether strict scrutiny or deferential review is the standard of review for equal protection challenges to governmental racial classification in prison.[22]

II. STRICT SCRUTINY IS THE STANDARD OF REVIEW FOR RACIAL CLASSIFICATION IMPOSED BY GOVERNMENT OUTSIDE THE PRISON CONTEXT

A court of review analyzes all governmental racial classification outside the prison context under strict scrutiny to "smoke out" illegitimate uses of race.[23] Under this standard, the government must prove any racial classifications it imposes "are narrowly tailored measures that further compelling governmental interests."[24] In *Grutter v. Bollinger*, for example, the United States Supreme Court used strict scrutiny to review the University of Michigan Law School's admissions policy.[25] The Law School narrowly tailored its policy, the first prong of the strict scrutiny analysis, by using racial classification in conjunction with several other criteria.[26] The Law School believed the policy was crucial to the school's mission to have a diverse student body.[27] In *Grutter*, the Court deferred to the Law School's judgment and held that their policy furthered a compelling government interest, thus satisfying both prongs of the strict scrutiny analysis.[28] Thus, the Supreme Court determined that the Law School could consider race in their candidate analysis because race was only one of several factors used.[29]

Three general propositions characterize the review of race-based governmental policies under strict scrutiny: skepticism, congruence, and

---

20. *Id.* at 506 (quoting Richmond v. J.A. Croson Co., 488 U.S. 469, 493 (1989) (plurality opinion)).

21. *Id.* at 500.

22. *Id.* at 502.

23. *See J.A. Croson Co.*, 488 U.S. at 493; *see also* 16B AM. JUR. 2D *Constitutional Law* § 815 (1998) (providing an overview of the strict scrutiny test).

24. *See, e.g.*, Adarand Constructors, Inc. v. Pena, 515 U.S. 200, 227 (1995) (holding that strict scrutiny is the appropriate standard of review for race-based government subcontractor preferences); *see also* Grutter v. Bollinger, 539 U.S. 306, 326 (2003) (holding that strict scrutiny should be used to examine a law school admissions policy that considers the race of applicants).

25. *Grutter*, 539 U.S. at 326.

26. *Id.* at 334.

27. *Id.* at 328.

28. *Id.* ("[W]e hold that the Law School has a compelling interest in attaining a diverse student body. The Law School's educational judgment that such diversity is essential to its educational mission is one to which we defer.").

29. *Id.* at 334-35.

consistency.[30] First, courts review racial classifications by the government under strict scrutiny with skepticism.[31] A race-based governmental policy raises skepticism because it determines the treatment of an individual according to a group classification.[32] Courts suspiciously inspect whether such group-based treatment by the government impinges on the individual's Fourteenth Amendment's guarantee of equal protection to the individual.[33]

The second characteristic of strict scrutiny is that it reconciles incongruence between equal protection provided by the Fifth Amendment and the Fourteenth Amendment.[34] In the 1940's, United States Supreme Court decisions implied that equal protection from federal racial classification afforded by the Fifth Amendment's Due Process Clause is less expansive than the equal protection from state racial classification guaranteed under the Fourteenth Amendment.[35] Though the Fifth Amendment, which applies to the federal government, contains no explicit Equal Protection Clause, the Supreme Court has applied

---

30. *See* Adarand Constructors, Inc. v. Pena, 515 U.S. 200, 223-24 (1995).

31. *Id.*; Fullilove v. Klutznick, 448 U.S. 448, 523 (1980) (Stewart, J., dissenting) ("[A]ny official action that treats a person differently on account of his race or ethnic origin is inherently suspect . . . .").

32. *Adarand*, 515 U.S. at 227.

> [T]he Fifth and Fourteenth Amendments to the Constitution protect *persons*, not *groups*. It follows from that principle that all governmental action based on race—a *group* classification long recognized as "in most circumstances irrelevant and therefore prohibited" should be subjected to detailed judicial inquiry to ensure that the *personal* right to equal protection of the laws has not been infringed.

Id. (quoting Hirabayashi v. United States, 320 U.S. 81, 100 (1943)).

33. *See id.*

34. *See* United States v. Paradise, 480 U.S. 149, 166, n.16 (1987) (plurality opinion of Brennan, J.) ("[T]he reach of the equal protection guarantee of the Fifth Amendment is coextensive with that of the Fourteenth [Amendment]"). *Compare* U.S. CONST. amend. V ("No person shall . . . be deprived of life, liberty, or property, without due process of law"), *and Adarand*, 515 U.S. at 213 ("Although this Court has always understood that [Fifth Amendment] Clause to provide some measure of protection against *arbitrary* treatment by the Federal Government, it is not as explicit a guarantee of *equal* treatment as the Fourteenth Amendment. . . .") (emphasis in original), *with* U.S. CONST. amend. XIV, § 1 ("No *State* shall . . . 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.") (emphasis added).

35. *See* Korematsu v. United States, 323 U.S. 214, 216 (1944) (holding that though racial classifications are "immediately suspect" and should be subjected to "the most rigid scrutiny," and concluding that the federal government's policy restricting Japanese-Americans from entering particular geographic areas was constitutional and within Congress's authority); *Hirabayashi*, 320 U.S. at 113 (Murphy, J., concurring) (approving of the majority's holding that racial classifications by the federal government are limited by the Fifth Amendment only to the extent that the classification would deny due process, Murphy added, "Modern war does not always wait for the observance of procedural requirements that are considered essential and appropriate under normal conditions. Accordingly I think the military arm, confronted with the peril of imminent enemy attack . . . made an allowable judgment at the time . . . ."). *See generally Adarand*, 515 U.S. at 213-14 (providing an analysis of the Supreme Court's equal protection jurisprudence deriving from the Fifth and Fourteenth Amendments in the 1940's).

the state Fourteenth Amendment equal protection requirement to the federal government.[36] The Court in *Hirabayashi v. United States* reasoned that the Fifth Amendment only affords protection from federal race-based policies that violate due process.[37] In *Bolling v. Sharpe*,[38] however, the Court found that the Fifth Amendment's protection against federal violations of due process and the Fourteenth Amendment's protection against state violations of equal protection provide equivalent safeguards against discriminatory racial classification.[39] The Court reasoned that both measures of protection originate from "our American ideal of fairness."[40] Thereafter, courts began to treat the equal protection requirements imposed by the Fifth and Fourteenth Amendments as "indistinguishable."[41] Regardless of whether the constitutional challenge originated under federal or state law, all *burdensome* racial classifications would be subject to strict scrutiny, providing congruence to this area of the law.[42]

Third, strict scrutiny provides consistent review of race-based governmental policies, regardless of whether the policies benefit or burden certain racial groups.[43] Through a single standard, government officials who craft race-based governmental policies face a certain and rigorous constitutional hurdle.[44] Until the 1990's, while the United States Supreme Court had set forth the general rule that classifications burdening racial groups would be subject to "the most rigid scrutiny," the Court had not adopted a uniform standard for government action intended to benefit certain groups.[45] In *Metro Broadcasting, Inc. v. FCC*,[46] the United States Supreme Court created a large caveat for strict scrutiny of governmental racial classification by holding that intermediate scrutiny[47] is the standard of review for "benign" federal racial classification.[48]

---

36. See *Hirabayashi*, 320 U.S. at 100.

37. *Id.*; *see also Adarand*, 515 U.S. at 213-14 (discussing the evolution of the Court's interpretation of the Equal Protection Clause as it applies to the federal government).

38. 347 U.S. 497 (1954).

39. *See id.* at 499 (finding "that the Constitution of the United States, in its present form, forbids, so far as civil and political rights are concerned, discrimination by the General Government, or by the States, against any citizen because of his race") (quoting Gibson v. Mississippi, 162 U.S. 565, 591 (1896)).

40. *Id.*

41. *See Adarand*, 515 U.S. at 217.

42. United States v. Paradise, 480 U.S. 149, 166, n.16 (1987) (plurality opinion of Brennan, J.) ("[T]he reach of the equal protection guarantee of the Fifth Amendment is coextensive with that of the Fourteenth . . . ."); *cf.* Weinberger v. Wiesenfeld, 420 U.S. 636, 638 n.2 (1975) ("This Court's approach to Fifth Amendment equal protection claims has always been precisely the same as to equal protection claims under the Fourteenth Amendment.").

43. *See* Richmond v. J.A. Croson Co., 488 U.S. 469, 493-94 (1989) (plurality opinion).

44. *Id.* at 493 ("[S]trict scrutiny . . . assur[es] that the legislative body is pursuing a goal important enough to warrant use of a highly suspect tool.").

45. *See Adarand*, 515 U.S. at 218-26.

46. 497 U.S. 547 (1990).

47. Policies pass intermediate scrutiny if they "serve important governmental objectives" and are "substantially related to achievement of those objectives." *Id.* at 565.

48. *Id.* at 565-67 (finding FCC racial policies that "enhanc[e] broadcast diversity" were

Despite recently finding strict scrutiny to be the standard of review for benign state racial classification in *Richmond v. J.A. Croson Company*,[49] the Court in *Metro Broadcasting* found that a federal governmental program that benefits a minority group, such as prioritizing the sale of radio and television stations to minorities, should be reviewed under an intermediate scrutiny standard.[50] *Metro Broadcasting* departed from prior cases in two major respects: First, the decision implied that courts can determine the character of a racial classification as burdensome or beneficial before assigning a standard of review.[51] Second, *Metro Broadcasting* refuted the congruence between federal and state racial classification, thus potentially reestablishing distinct equal protection analyses under the Fifth and Fourteenth Amendments.[52]

In *Adarand Constructors, Inc. v. Pena*, the Court rejected the *Metro Broadcasting* exception to strict scrutiny for race-based governmental policies that benefit racial groups.[53] While *Metro Broadcasting* had asserted strict scrutiny failed to take into account "relevant differences" between beneficial and burdensome race-based policies,[54] *Adarand* claimed that "strict scrutiny *does* take 'relevant differences' into account—indeed, that is its fundamental

---

"benign"). *But see Adarand*, 515 U.S. at 225 (overruling *Metro Broadcasting* and criticizing its failure to "explain how to tell whether a racial classification should be deemed benign," other than by examining its legislative scheme and history) (citation omitted).

49. 488 U.S. 469, 493-94 (1989) (holding that strict scrutiny is the standard of review for an ordinance benefitting minorities that required the city of Richmond to award thirty percent of construction contracts to minority contractors).

50. *Metro Broadcasting, Inc.*, 497 U.S. at 564-66.

51. *See Adarand*, 515 U.S. at 227. *But see J.A. Croson Co.*, 488 U.S. at 493 (stating "there is simply no way of determining what classifications are 'benign' or 'remedial' and what classifications are in fact motivated by illegitimate notions of racial inferiority or simple racial politics"). *See generally* Stephen L. Carter, *When Victims Happen To Be Black*, 97 YALE L.J. 420 (1988) (discussing how a race-conscious society makes well-intentioned policies that ultimately perpetuate categorization of particular races).

52. *See Metro Broadcasting, Inc.*, 497 U.S. at 564-66; *see also Adarand*, 515 U.S. 226-27 (finding that *Metro Broadcasting* undermined the skepticism and consistency set forth by earlier equal protection cases by rejecting the congruence between state and federal racial classifications).

53. *See Adarand*, 515 U.S. at 227 ("'[A] free people whose institutions are founded upon the doctrine of equality' should tolerate no retreat from the principle that government may treat people differently because of their race only for the most compelling reasons." (quoting Hirabayashi v. United States, 320 U.S. 81, 100 (1943))).

54. *Metro Broadcasting, Inc.*, 497 U.S. at 565; *see also Adarand*, 515 U.S. at 245 (Stevens, J., dissenting) (arguing that the Supreme Court should recognize a difference between beneficial and burdensome classifications). In his dissenting opinion, Justice Stevens illustrated,

> The consistency that the Court espouses would disregard the difference between a "No Trespassing" sign and a welcome mat. It would treat a Dixiecrat Senator's decision to vote against Thurgood Marshall's confirmation in order to keep African-Americans off the Supreme Court as on a par with President Johnson's evaluation of his nominee's race as a positive factor.

Id.

purpose."[55] Under strict scrutiny, courts theoretically learn "relevant" details from the policy creators or administrators, who may explain why a policy was originally implemented, what success the policy has achieved, and what the policy may achieve in the future.[56] The burden is on the government policymakers to explain precisely why the policy passes the narrow tailoring and furthering governmental interest requirements of the strict scrutiny standard.[57] Strict scrutiny affords courts an opportunity to carefully examine the legitimacy of racial classification in cases where the beneficial or burdensome nature of a policy is not clear.[58] In *Adarand*, the Court found that strict scrutiny should be used to examine *all* governmental policies using racial classifications.[59] By overruling *Metro Broadcasting* in *Adarand*, the Court virtually shut the door on exceptions to strict scrutiny of race-based policies imposed by the state and federal government.[60]

III. DEFERENTIAL REVIEW FOR PRISON POLICIES THAT IMPLICATE CONSTITUTIONAL RIGHTS

In *Turner v. Safley*,[61] the United States Supreme Court held that prison administrators' policies may impinge on prisoners' constitutional rights only if the policies reasonably relate to legitimate penological interests.[62] After *Turner*, the court continued to apply a reasonable relationship test when reviewing inmates' constitutional claims.[63] In *Shaw v. Murphy*, the United States Supreme Court noted that in applying the standard set forth in *Turner*, the Court effectively adopted "a unitary, deferential standard for reviewing prisoners' constitutional claims."[64] Consequently, in *Shaw*, the Court upheld the constitutional protections that prisoners had enjoyed since the 1960's,[65] but

---

55. *Adarand*, 515 U.S. at 228 (finding that when courts apply a rigorous standard of review, advocates of a race-based policy must make clear to the court how all aspects of the policy serve compelling government interests and are narrowly tailored, thereby displaying important characteristics of the policy that the court may find beneficial or burdensome to affected groups).

56. *Id.* at 227-28.

57. *Id.*

58. *Id.* at 228, 236.

59. *See id.* at 236.

60. *Id.*

61. 482 U.S. 78 (1987).

62. *See id.* at 89, 91 (holding that a restriction on inmates' right to marry inmates in different prisons did not reasonably relate to penological interests, such as inmate and staff security, but a restriction on inmates' correspondence with inmates in different prisons did reasonably relate to issues of security).

63. *See, e.g.*, Shaw v. Murphy, 532 U.S. 223, 229 (2001).

64. *Id.*

65. Prisoners enjoy some constitutional protection in recent decades, despite a long history of lost constitutional rights during incarceration. *See generally* Wolff v. McDonnell, 418 U.S. 539 (1974) (holding prisoners have certain "minimal" due process rights); Lee v. Washington, 390 U.S. 333 (1968) (*per curiam*) (holding prisoners have the right to be free from

the Court also reaffirmed the traditional power of prison administrators to limit those rights without strict judicial oversight.[66]

Before the Court adopted a standard of review, federal courts had "expressly reserved the question of the proper standard of review to apply in cases involving questions of prisoners' rights."[67] For example, in *Procunier v. Martinez*, the Supreme Court refused to resolve "broad questions of 'prisoners'' rights," instead adopting a narrow holding limited to the specific facts of the case.[68] As a result, a court could choose to review a prisoner's constitutional challenge under a strict scrutiny standard or a deferential standard, depending upon the facts of the case.[69] The Court's refusal in *Martinez* to adopt a standard of review reflected the federal court system's broad, "hands-off" philosophy regarding prison administration.[70] The Supreme Court's failure to adopt a standard of review resulted in multiple federal court standards, which generated inconsistent constitutional rulings.[71]

In *Pell v. Procunier*,[72] the Court departed from the "hands-off"[73] philosophy and set forth a deferential standard of review for prisoners' constitutional challenges.[74] Although the Court intended for this new standard to bring consistency to federal rulings on prisoners' constitutional challenges, the Court also intended to limit federal court scrutiny of prison policies.[75] In *Jones v. North Carolina Prisoners' Union*,[76] the Court further embraced

---

racial discrimination).

66. *See Shaw*, 532 U.S. at 230-31 (rejecting additional federal court oversight of prisoners' constitutional rights as inconsistent with *Turner* analysis); *Turner*, 482 U.S. at 89.

67. *Turner*, 482 U.S. at 85-86 (quoting Procunier v. Martinez, 416 U.S. 396, 409 (1974)) (internal quotation marks omitted).

68. *Martinez*, 416 U.S. at 408.

69. *See id.* at 406-07 (comparing standards of review for "constitutional challenges to censorship of prisoner mail").

70. *Id.* at 404-05. The federal courts historically adopted a "hands-off" philosophy toward prison administration issues because prison policies required expert planning and distribution of resources which are "peculiarly within the province of the legislative and executive branches of government." *Id.* at 405.

71. *See id.* at 406 ("This Court has not previously addressed this [standard of review] question, and the tension between the traditional policy of judicial restraint regarding prisoner complaints and the need to protect constitutional rights has led the federal courts to adopt a variety of widely inconsistent approaches to the problem.").

72. 417 U.S. 817, 825 (1974) (holding that a prison policy prohibiting media interviews with certain prisoners did not violate the First Amendment because all inmates could interact with media through intermediaries like family, clergy, and attorneys who visited prisons).

73. *Martinez*, 416 U.S. at 404.

74. *Pell*, 417 U.S. at 827 (finding courts should ordinarily defer to wardens' expert judgment).

75. *Id.* (finding that policies which affect penological interests like prison security are "peculiarly within the province and professional expertise of corrections officials, and, in the absence of substantial evidence in the record to indicate that the officials have exaggerated their response to these considerations, courts should ordinarily defer to their expert judgment in such matters").

76. 433 U.S. 119 (1977).

deferential review by demanding that prisoners carry the burden of proving policies are "exaggerated," "unreasonable" responses to penological interests.[77] The Court's holding in *Jones*, that deference must end when policies become unreasonable, set the stage for the reasonable relationship test outlined in *Turner*.[78]

The Court set forth the reasonable relationship test, also called the *Turner* Standard, to determine whether a penological interest justifies burdening an inmate's constitutional rights.[79] Under the test, if a policy impinges on a constitutional right retained by the prisoner upon incarceration,[80] then the policy may only burden that right if it reasonably relates to a legitimate penological interest.[81] Four factors help determine the reasonableness of a policy burdening constitutional rights: a policy's reasonable relation to penological interests,[82] whether the prisoner has an alternative means of exercising a burdened right,[83] the "ripple effect" caused by accommodation of an asserted right,[84] and the existence of alternatives to the policy.[85]

The reasonable relationship test allows a warden to relate a policy that impinges on a prisoner's constitutional right to the broad penological interests that wardens themselves often define.[86] The Court adopted the reasonable relationship test because strict scrutiny "would seriously hamper [prison officials'] ability to anticipate security problems and to adopt innovative

---

77. *Id.* at 128-29 (holding that the warden did not violate inmates' free speech rights by refusing to deliver prisoner mail that solicited membership in a prisoner labor union because preventing formation of the union was reasonably related to the penological interest of order and security).

78. *See id.* at 127-28.

79. *See* Turner v. Safley, 482 U.S. 78, 89 (1987).

80. An inmate "retains those [constitutional] rights that are not inconsistent with his status as a prisoner or with the legitimate penological objectives of the corrections system." *Pell*, 417 U.S. at 822.

81. *Turner*, 482 U.S. at 89.

82. *See id.* at 89-90 ("Thus, a regulation cannot be sustained where the logical connection between the regulation and the asserted goal is so remote as to render the policy arbitrary or irrational.").

83. *Id.* at 90; *see also Pell*, 417 U.S. at 825 (1974) (finding that restrictions on inmate interviews with media did not violate the First Amendment because reasonable and effective alternative means of communication with the media remained open through intermediaries such as family and attorneys).

84. *Turner*, 482 U.S. at 90 ("When accommodation of an asserted right will have a significant 'ripple effect' on [distribution of resources like guards, supplies, and administrators' time], courts should be particularly deferential to the informed discretion of corrections officials.").

85. *See id.* at 90-91. When an alternative policy exists and requires little effort to implement, then that alternative policy may be evidence that the current policy is an "exaggerated response." *Id.* at 90. This factor does not set forth a "least restrictive alternative" test that demands prison officials "set up and then shoot down every conceivable alternative method of accommodating the [prisoner's rights]." *Id.* at 90-91.

86. *See id.* at 89.

solutions to the intractable problems of prison administration."[87] In *Washington v. Harper*, the Court rejected any exception to the *Turner* Standard in circumstances where the needs of prison administration implicate constitutional rights.[88]

IV. JOHNSON V. CALIFORNIA: STRICT SCRUTINY IS THE STANDARD OF REVIEW FOR RACIAL CLASSIFICATION IN PRISON

The United States Supreme Court settled the debate over which standard of review applies to race-based prison policy by reaching a 5-3 decision in *Johnson v. California*.[89] Justice O'Connor delivered the opinion of the Court, joined by Justices Kennedy, Souter, Ginsburg, and Breyer.[90] Chief Justice Rehnquist took no part in the decision.[91] The Court held that strict scrutiny is the standard of review for prisoner challenges to race-based policies burdening the constitutional right to equal protection under the Fourteenth Amendment.[92]

The majority opinion first explained why strict scrutiny is the Court's best tool for "smok[ing] out" the primary danger of governmental racial classification—racial discrimination.[93] Strict scrutiny places the burden of proving constitutionality upon the government.[94] The majority found that the government must show that it has narrowly tailored a policy that uses racial classifications and that the policy serves a compelling government interest.[95] As the government seeks to pass this two-part test, courts can uncover illegitimate uses of race involving discriminatory motives.[96]

The *Johnson* majority contended that the heightened standard used in *Lee v. Washington* to review racial segregation in prison closely resembled strict scrutiny, rather than deferential review.[97] In *Lee*, the Court affirmed a three-judge court's decision to strike down racial segregation in Alabama prisons.[98] The majority opinion in *Johnson* noted that a prison segregation policy could only pass *Lee*'s heightened standard "in particularized circumstances" that "take into account racial tensions in maintaining security, discipline, and good

87. *Id.*

88. Washington v. Harper, 494 U.S. 210, 224 (1990).

89. Johnson v. California, 543 U.S. 499, 501-02 (2005).

90. *Id.* at 501.

91. *Id.*

92. *Id.* at 515.

93. *Id.* at 506 (quoting Richmond v. J.A. Croson Co., 488 U.S. 469, 493 (1989) (plurality opinion)).

94. *Id.* at 506 & n.1 ("We put the burden on state actors to demonstrate that their race-based policies are justified.").

95. *Id.* at 509.

96. *Id.* at 506 (finding a "searching judicial inquiry" important for race-based measures (quoting *J.A. Croson Co.*, 488 U.S. at 493)).

97. *Id.* at 506-07. The Court in *Lee* never called its standard of review "heightened." *See id.*; Lee v. Washington, 390 U.S. 333, 333-34 (1968) (per curiam).

98. *Lee*, 390 U.S. at 333.

order in prisons and jails." [99] The *Johnson* majority found this heightened standard of review in *Lee* to be consistent with strict scrutiny because both standards reject racial segregation in prison unless the policy is narrowly applied[100] and serves important government interests.[101] Additionally, the *Johnson* majority found the review of racial classification in *Lee* similar to their own review because *Lee* sought no "dilution of this Court's firm commitment to the Fourteenth Amendment's prohibition of racial discrimination."[102]

The majority rejected the defendants' contention that their policy "'equally' segregated" races, and was therefore exempt from strict scrutiny.[103] The defendants argued that temporary segregation in reception centers burdens or benefits races equally because the defendants segregate all prisoners.[104] The defendants contended that although their policy separates the races from each other for the evaluative period, all races enjoy equal treatment within those separated environments during that time.[105] The Court, citing *Brown v. Board of Education*,[106] rejected the argument that segregation can ever produce equal treatment.[107] Instead, racial segregation can produce the illusion of equal treatment that frequently results in administrators and other prisoners wrongfully assigning qualities or stereotypes to cells or cellblocks based upon race.[108] This process stigmatizes inmates and may "incite racial hostility,"[109] especially when tension originating from racial gangs already runs high.[110] Consequently, the *Johnson* majority "refuse[d] to resurrect" what it already rejected in *Brown*.[111] The Court reiterated that equally segregating people by race does not create equal treatment under the Fourteenth Amendment.[112]

In holding that the defendants' race-based policy deserved strict scrutiny, the Court also noted that the policy may be ineffective; prison experts have

---

99. *Johnson*, 543 U.S. at 507 (quoting *Lee*, 390 U.S. at 334).

100. *Id.* (finding narrow tailoring demanded by strict scrutiny and "particularized circumstances" are roughly equivalent (quoting *Lee*, 390 U.S. at 334)).

101. The Court finds that the use of racial classification for "necessities of prison security and discipline" to be equivalent to compelling government interests required by strict scrutiny. *Id.* at 512 (quoting *Lee*, 390 U.S. at 334).

102. *Id.* at 507 (quoting *Lee*, 390 U.S. at 334).

103. *Id.* at 506.

104. *See id.*

105. *See id.*

106. 347 U.S. 483 (1954).

107. *Johnson*, 543 U.S. at 506 (citing *Brown*, 347 U.S. at 495).

108. *See id.*; *see also* Shaw v. Reno, 509 U.S. 630, 643 (1993) (finding racial classifications "stigmatize individuals by reason of their membership in a racial group").

109. *Shaw*, 509 U.S. at 643.

110. *See Johnson*, 543 U.S. at 502 (identifying "Mexican Mafia, Nuestra Familia, Black Guerilla Family, Aryan Brotherhood, and [the] Nazi Low Riders" as five major prison gangs whose cultures are "violent and murderous").

111. *Id.* at 506.

112. *Johnson*, 543 U.S. at 506; *see also Brown*, 347 U.S. at 495 (holding that the segregated plaintiffs were "deprived of the equal protection of the laws guaranteed by the Fourteenth Amendment").

found that racial segregation actually exacerbates racial violence.[113] The Court noted that "the Federal Bureau of Prisons (BOP) expressly prohibit[s] racial segregation."[114] In an *Amicus Curiae* brief to the United States Supreme Court, the BOP asserted that integration actually mitigates interracial violence and "prepares inmates for re-entry into society."[115] In addition, the Court found it particularly significant that no other state uses race as the predominant factor in cell assignments.[116] Former state correctional officials from six states asserted that a "blanket policy" like temporary racial segregation is "contrary to sound prison management."[117] Moreover, a sociological study provided evidence to the Court that the prison policy may, in fact, increase the rate of interracial violence.[118] The study showed reduced violence in racially integrated cells over a period of ten years.[119] Because racial classification is "immediately suspect" in all contexts, and the defendants' race-based policy is outside the mainstream of prison policy, the Court found the race-based policy questionable and in need of strict scrutiny.[120]

After the majority demonstrated why strict scrutiny should be the standard of review for race-based prison policies, the Court explained why the *Turner* standard of deferential review should not apply.[121] The majority first made clear that the Court has never applied *Turner* to equal protection challenges arising from racial classification.[122] Instead, the Court has applied *Turner* to rights "inconsistent with proper incarceration."[123] The Court explained that

---

113. *See Johnson*, 543 U.S. at 507-09.

114. *Id.* at 508.

115. *Id.* at 509.

116. *Id.* at 508 & n.2.

117. *Id.* at 508 (quoting a Brief of Amici Curiae for Former State Corrections Officials at 19, Johnson v. California, 543 U.S. 499 (2005) (No. 03-636)). The *amici* acknowledge that inmates in Texas and Oklahoma reception centers "are not generally assigned randomly to racially integrated cells," but the inmates are not totally precluded from integrated cells if they seek them. *Id.* at 508 n.2 (citing Oklahoma Dept. of Corrections, Policies and Procedures, Operations Memorandum No. OP-030102, Inmate Housing (Sept. 16, 2004), *available at* http://www.doc.state.ok.us/Offtech/ op030102.htm; Texas Dept. of Criminal Justice, Security Memorandum No. SM-01.28, Assignment to General Population Two-Person Cells (June 15, 2002)). The Court cites the Brief for Former State Corrections Officials that states, "To the extent that race is considered in the assignment calculus in Oklahoma, it appears to be one factor among many . . . ." *Johnson*, 543 U.S. at 508 n.2 (quoting Brief, *supra*, at 20 n.10). The Court does not conclude that Texas and Oklahoma actually have a multi-factor calculus, or that this potential calculus that classifies inmates by race would pass strict scrutiny. *Id.*

118. *Johnson*, 543 U.S. at 507-08 (citing Chad Trulson & James W. Marquart, *The Caged Melting Pot: Toward an Understanding of the Consequences of Desegregation in Prisons*, 36 LAW & SOC'Y REV. 743, 774 (2002)).

119. *Id.*

120. *See id.* at 509 (quoting Shaw v. Reno, 509 U.S. 630, 642 (1993)).

121. *Id.* at 509-13.

122. *Id.* at 510.

123. *Id.* (noting that "a prison inmate retains those First Amendment rights that are not inconsistent with his status as a prisoner" (quoting Pell v. Procunier, 417 U.S. 817, 822 (1974)). One simple example of a right inconsistent with proper incarceration is the right to bear arms.

Turner applies to the constitutional rights necessarily limited by incarceration, such as the right to freely associate.[124] The majority found that penological interests do not justify impinging upon the right to equal protection under the Fourteenth Amendment.[125] Because the defendants' penological interests are not generally served by restricting an inmate's right to racial integration, the majority found that the plaintiff deserves strong constitutional protection afforded by strict scrutiny.[126]

The majority also rejected Justice Thomas's contention that the reasonable relationship test adequately filters out race-based policies motivated by racial discrimination.[127] The majority argued the reasonable relationship test ineffectively regulates race-based policies for two reasons.[128] First, prison officials have no duty to choose a policy that implicates Fourteenth Amendment protection when both racial and non-racial policies exist and each serves the same purpose, like heightened security.[129] Second, race-based prison policies may pass the reasonable relationship test if their intended purpose reasonably relates to a penological interest; however, passing the test does not guarantee that the purpose is achieved.[130] Therefore, if a prison administrator believes that a race-based policy like temporary racial segregation prevents violence, when in fact it contributes to more violence, the policy still passes the reasonable relationship test and may be implemented.[131] Even if the policy fails to achieve its intended purpose, administrators have no duty to eliminate the policy.[132]

The majority ended its analysis by asserting that the defendants' policy would not necessarily fail strict scrutiny upon remand.[133] The Court refuted the defendants' contentions that strict scrutiny "handcuff[s] prison administrators"[134] and is "strict in theory, but fatal in fact."[135] Referencing an example of a policy passing strict scrutiny, the majority cited *Grutter v.*

---

124. *Id.* (citing Overton v. Bazzetta, 539 U.S. 126, 131 (2003)) (finding restrictions upon inmates' rights to association were reasonably related to penological interest of security under the *Turner* standard).

125. *Id.* at 510-11.

126. *Id.* at 510-12.

127. *Id.* at 513-14.

128. *See id.*

129. *Id.* at 513 (noting that the Court in *Turner* explicitly stated the test was not a "least restrictive alternative" test).

130. *Id.*

131. *Id.* (finding that in their application of *Turner*, the appellate court "did not have to agree that the policy actually advances the CDC's legitimate interest, but only [that] 'defendants might reasonably have thought that the policy would advance its interests'" (quoting Johnson v. California, 321 F.3d 791, 803 (2003))).

132. *Id.*

133. *Id.* at 515 ("The fact that strict scrutiny applies 'says nothing about the ultimate validity of any particular law; that determination is the job of the court applying strict scrutiny.'" (quoting Adarand Constructors, Inc. v. Pena, 515 U.S. 200, 230 (1995))).

134. *Id.* at 514.

135. *Id.* (quoting *Adarand*, 515 U.S. at 237) (internal quotation marks omitted).

Bollinger, in which the University of Michigan Law School's race-based admissions policy passed the two-part strict scrutiny test.[136] The *Johnson* majority indicated that the defendants' race-based policy may indeed be necessary to prevent interracial prison violence, which promotes a "compelling government interest."[137] The Court also emphasized that the defendants must show on remand how they narrowly tailored their policy to prevent interracial violence.[138] Without narrow tailoring, the Court clarified that the defendants' policy of assigning cellmates on the basis of race will fail strict scrutiny.[139]

The Supreme Court reversed the decision of the Court of Appeals for the Ninth Circuit and remanded the case to the lower court.[140] For a lower court to find the defendants' policy constitutional upon remand, the government must carry the burden of showing its policy is a narrowly tailored measure that furthers compelling governmental interests.[141]

Justice Ginsburg, in a concurring opinion joined by Justice Souter and Justice Breyer,[142] agreed that the Court should review governmental racial classification under strict scrutiny when a government policy burdens constitutional rights.[143] Justice Ginsburg, however, reaffirmed a distinction she drew in *Grutter v. Bollinger* between the standard of review for policies that burden constitutional rights and policies that benefit constitutional rights.[144] Justice Ginsburg rejected the idea that strict scrutiny is the appropriate standard of review for race-based policies that *benefit* minority racial groups.[145] Because Justice Ginsburg found that defendants' policy did not involve a remedial race-based policy, there was no basis for objection in *Johnson*.[146]

Justice Stevens dissented from the majority's decision to remand the case, claiming that the current record contained adequate proof for this Court to determine that the policy is unconstitutional.[147] Justice Stevens argued that under either the reasonable relationship test or strict scrutiny, the Court should rule that the CDC blunderbuss policy violates the Equal Protection Clause.[148]

---

136. *Id.* In *Grutter*, the University of Michigan Law School had a compelling interest in attaining a diverse student body, and the University narrowly tailored its admissions program to serve a compelling interest because it used race as one of many factors in candidate analysis. Grutter v. Bollinger, 539 U.S. 306, 328 (2003).

137. *Johnson*, 543 U.S. at 512.

138. *Id.* at 515.

139. *Id.*

140. *Id.*

141. *Id.*

142. *Id.* at 516 (Ginsburg, J., concurring).

143. *Id.*

144. *Id.* (asserting remedial race-based policies should be reviewed under less strict standard of review than strict scrutiny); Grutter v. Bollinger, 539 U.S. 306, 344 (2003) (Ginsburg, J., concurring).

145. *Johnson*, 543 U.S. at 516 (Ginsburg, J., concurring).

146. *Id.*

147. *Id.* at 517 (Stevens, J., dissenting).

148. *Id.* at 517-18, 521 (asserting the fact that a policy was unwritten reflects "the evident lack of deliberation that preceded its creation"). Justice Stevens found that no evidence

The Court had previously acknowledged that its ruling on a standard of review "alters the playing field" for the defendants and necessitates new preparation before a proper rehearing on remand.[149] Disagreeing with the majority's decision to remand, Justice Stevens believed that the defendants had every opportunity to prove that their policy passed both a deferential standard and strict scrutiny, but they failed to do either.[150]

Justice Stevens attacked the defendants policy of assigning cellmates on the basis of race by arguing for individualized assessment of each prisoner.[151] "The policy currently uses an inmates race [as] a proxy for gang membership, and gang membership [as] a proxy for violence."[152] CDC officials compile records of individual inmates to learn their gang affiliations, criminal records, and general compatibility with other inmates.[153] Justice Stevens contended that the CDC could circumvent automatic segregation of a new or transferred inmate if the reception center received pre-sentence inmate reports upon inmate arrival.[154] Consequently, at the outset of incarceration, the prison officials could assign reception center cellmates on the basis of each inmate's individual file instead of his race.[155]

Justice Stevens disapproved of a policy that implements a sixty day evaluative period for transferees, which fosters repetitive racial segregation.[156] He noted that transferees represent eighty-five percent of the inmates subject to the reception center cellmate policy.[157] Justice Stevens reasoned that after prison officials evaluated an inmate during an initial intake, it was redundant to examine the prisoners and subject them to racial segregation again.[158] Justice Stevens suspected that the cell assignment policy was an "exaggerated response" to administrative inefficiency.[159] Because Justice Stevens found the CDC failed to make a serious, good faith consideration of workable race-neutral alternatives[160] even when obvious, easy alternatives existed,[161] he stated that the policy was unconstitutional under any standard of review.[162]

---

supported the defendants' contention that the race-based policy saves lives. *Id.* at 520. He noted that the defendants have never experimented with other policies to determine whether their segregation policy actually reduces the rates of interracial violence. *Id.* at 521 ("[O]ne might assume that the CDC came to its policy only as a last resort."). Therefore, he found no reasonable relationship between the policy and its goal of reducing violence. *See id.* at 523, n.3.

149. Adarand Constructors, Inc. v. Pena, 515 U.S. 200, 237 (1995).
150. *Johnson*, 543 U.S. at 517 (Stevens, J., dissenting).
151. *Id.* at 521.
152. *Id.* at 517.
153. *Id.* at 521-22.
154. *See id.* at 522.
155. *Id.*
156. *Id.*
157. *Id.* at 521.
158. *Id.*
159. *See id.* at 523 n.3 (quoting Turner v. Safley, 482 U.S. 78, 90 (1987)).
160. *Id.* at 522 (quoting Grutter v. Bollinger, 539 U.S. 306, 339 (2003)).
161. *Id.* at 523 (quoting *Turner*, 482 U.S. at 90).
162. *Id.*

Justice Thomas, joined by Justice Scalia, also dissented from the majority.[163] Justice Thomas argued that the Court should apply a deferential standard to racial classification in prisons because the defendants, who are expert prison managers, believe their policy reasonably relates to reducing violence and saving lives.[164] Justice Thomas began his dissent by recognizing the clashing standards of review set forth by the Court in *Harper* and *Adarand*.[165] In *Harper*, the Court held that "the [relaxed] standard of review we adopted in *Turner* applies to *all* circumstances in which the needs of prison administration implicate constitutional rights."[166] In *Adarand*, however, the Court held that "*all* racial classifications reviewable under the Equal Protection Clause must be strictly scrutinized."[167] The *Johnson* majority found that *Turner* did not apply to the facts of *Johnson* because prison administrators did not *need* to implicate inmates Fourteenth Amendment rights to serve penological interests.[168] Justice Thomas found that both holdings applied to the facts of *Johnson*, but he argued the deferential standard was appropriate because it affords deference to difficult decisions that may save lives in a hostile environment.[169]

In addition, Justice Thomas criticized the majority's finding that *Turner* does not apply to the constitutional right of equal protection from racial discrimination under the Fourteenth Amendment.[170] The majority found that the *Turner* reasonable relationship test applies only to restrictions on rights inconsistent with proper incarceration.[171] Justice Thomas argued that the majority's analysis requires two determinations by a court: a determination of whether a right is "inconsistent with proper incarceration" and a standard of review determination.[172] Justice Thomas rejected this approach because it, contrary to tradition, deprives the prison administrators from conclusively deciding what rights are inconsistent with incarceration.[173] The problem with the majority's analysis, Justice Thomas noted, is that while courts determine what rights incarceration implicates, prison officials have a special knowledge

---

163. *Id.* at 524-50 (Thomas, J., dissenting).

164. *Id.* at 524.

165. *Id.* (citing holdings from Adarand Constructors, Inc. v. Pena, 515 U.S. 200 (1995) and Washington v. Harper, 494 U.S. 210 (1990)).

166. *Harper*, 494 U.S. at 224 (1990) (emphasis added).

167. *Adarand*, 515 U.S. at 224 (emphasis added).

168. *Johnson*, 543 U.S. at 510 (finding *Turner* applies *only* to rights that are "inconsistent with proper incarceration," which does not include the right to equal protection from racial discrimination (quoting Overton v. Bazzetta, 539 U.S. 126, 131 (2003))).

169. *Id.* at 524 (Thomas, J., dissenting) ("the majority is concerned with sparing inmates the indignity and stigma of racial discrimination. California is concerned with their safety and saving their lives.") (citation omitted).

170. *See id.* at 531.

171. *Id.* at 510 (majority opinion).

172. *Id.* at 541-42 (Thomas, J., dissenting) (questioning whether court "has some implicit notion of what a proper prison ought to look like and how it ought to be administered").

173. *Id.* at 542.

and expertise in the area.[174] Justice Thomas asserted that penological interests determine what rights should be impinged, and prison officials, as day-to-day managers, should determine the interests of the penal system.[175]

After Justice Thomas argued that a deferential standard applies to race-based prison policies, he analyzed the defendants' policy under the four factors of the reasonable relationship test.[176] First, he argued temporary segregation is reasonably related to the penological interest of prison security.[177] Justice Thomas noted that prison security is a legitimate penological interest that is jeopardized by interracial violence in many California prisons.[178] Additionally, he found that the risk of interracial violence is especially high during the evaluative period in reception centers because staff are unaware of an inmate's gang affiliation or racist tendencies[179] and segregating unevaluated prisoners may minimize the risk of interracial violence.[180] Because temporary segregation may prevent interracial violence, the policy is reasonably related to prison security.[181]

Second, Justice Thomas briefly noted how prisoners have alternative means of exercising their restricted right to racially integrated life.[182] He noted that the racial segregation policy does not extend to spheres of life outside the double cells in reception centers.[183] For example, inmates are racially integrated activities in the dining hall, work and recreation areas.[184] Additionally, inmates have total control over integration in their lives once the evaluative period ends.[185] Justice Thomas found that the policy provided sufficiently alternative means for inmates to enjoy equal protection through racial integration in all other facets of prison life.[186]

Third, Justice Thomas found that accommodating the inmates' right to equal protection may drain administrative resources,[187] due to the high number

---

174. *Id.*

175. *Id.* at 541-42.

176. *Id.* at 534-38.

177. *Id.* at 534; *see also* Turner v. Safley, 482 U.S. 78, 91-93 (1987) (discussing the penological interest of prison security in the context of prohibiting inmate-to-inmate correspondence).

178. *Johnson*, 543 U.S. at 535 (Thomas, J., dissenting) (finding that the "racial component to prison violence is impossible for prison administrators to ignore").

179. *Id.*

180. *Id.*

181. *Id.*

182. *See id.* at 536; *see also Turner*, 482 U.S. at 90 (finding that an alternative method of expressing a restricted right is a factor in determining the reasonableness of a prison policy that impinges on a constitutional right).

183. *Johnson*, 543 U.S. at 536 (Thomas, J., dissenting).

184. *Id.*

185. *Id.*

186. *Id.*

187. *See id.* at 536-37 (finding that if courts do not defer to prison officials' policies, and this failure to defer causes administrative expense, then it is more likely that a reasonable relationship exists between the prison policy impinging constitutional rights and a legitimate

of inmates in prison and the limited number of staff available to oversee cells.[188] Justice Thomas hypothesized that interracial cells would necessitate more supervision and further burden prison guards.[189] Consequently, Justice Thomas concluded that the interracial cellmate policy may lead to overworked guards and may drain administrative resources.[190]

Fourth, Justice Thomas noted that Johnson had failed to show that no obvious, easy alternatives exist.[191] In his dissenting opinion, Justice Stevens suggested that officials could receive and process an inmate's pre-sentence report before he or she arrives and assign the inmate a cellmate based on the report, rather than upon race.[192] Justice Thomas refuted this alternative on the grounds that processing the reports would demand more of the already limited manpower and resources.[193] Additionally, Justice Thomas found that the reports "would not tell prison officials what they need to know" about an inmate's racially violent tendencies.[194]

Justice Thomas concluded his dissent with an analysis implying that the CDC policy passes strict scrutiny.[195] The plaintiff conceded that the government has a compelling government interest in implementing policies that further inmate and staff safety.[196] On remand, the only issue for the court will be whether the policy is narrowly tailored and Justice Thomas implied that the defendants will succeed in showing narrow tailoring.[197] Justice Thomas argued that the defendants actually use a multi-factor test in determining cell-mates, and hinted that this is evidence of a narrowly tailored use of race.[198] In sum, Justice Thomas dissented from the majority's decision that strict scrutiny should apply to the race-based prison policy, but he agreed with the majority's

---

penological interest); *see also Turner*, 482 U.S. at 90 (finding added expense of the accommodation of an asserted right is a factor in determining the reasonableness of the policy).

188. *Johnson*, 543 U.S. at 536 (Thomas, J., dissenting).

189. *Id.*

190. *See id.* at 536-37 (finding that drained human resources could lead to ineffective responses to violence).

191. *Id.* at 537; *see also Turner*, 482 U.S. at 90 (finding that if an alternative policy exists that does not impinge upon constitutional rights, then this is a factor considered in determining the reasonableness of the policy).

192. *Johnson*, 543 U.S. at 522 (Stevens, J., dissenting) (finding that an individualized assessment of new or transferred inmates is possible through efficient reporting practices).

193. *Id.* at 538 (Thomas, J., dissenting).

194. *Id.* at 537-38 (finding pre-sentence reports often fail to provide information about an inmate's racist tendencies or gang affiliation).

195. *Id.* at 547-50.

196. *Id.* at 54-48. The plaintiff also acknowledged that a policy that considered race as one factor in analysis of dispositions for violence would be constitutional. *Id.* at 535.

197. *Id.*

198. *Id.* (finding the policy neutral because it utilized a multi-factor test for determining cellmates based upon factors in addition to race such as geographic or national origin, age, physical size, mental health, medical needs, criminal history, and, of course, gang affiliation).

decision to remand the case so that the CDC could present evidence of narrow tailoring.[199]

V. IMPLICATIONS OF *JOHNSON V. CALIFORNIA*

The United States Supreme Court determined in *Johnson* that strict scrutiny is the standard of review for racial classification in prison despite a tradition of deferentially reviewing prison policies.[200] The Court cast the burden upon prison administrators to prove their housing policies are narrowly tailored and further compelling government interests.[201] The Court also reaffirmed that strict scrutiny applies to both burdensome *and* benign racial classification imposed by the government.[202] Upon remand, a lower court has the rare opportunity[203] to rule that strict scrutiny is not "fatal in fact" when applied to a racially burdensome governmental policy challenged on equal protection grounds.[204] The CDC housing policy may pass strict scrutiny if prison administrators select cellmates based on multi-factor, individualized assessments of new inmates and produce data showing an actual decrease in race-based gang violence.[205]

Upon remand of *Johnson* to a lower court, prison administrators in states like California, Texas and Oklahoma may argue that they narrowly tailor their race-based cellmate assignments by using multi-factor intake tests, in which race is only one factor along with others such as gang affiliation.[206] However, under *Johnson's* strict scrutiny standard, courts should look past the method by which cellmates are assigned to the actual results. If results indicate non-gang affiliated individuals are segregated on the basis of race, then a race-based

---

199. *Id.* at 524, 549.

200. *Id.* at 509, 515 (majority opinion).

201. *Id.* at 505.

202. *Id.*

203. *See* Korematsu v. United States, 323 U.S. 214 223-24 (1944) (upholding an order that excluded people of Japanese ancestry from particular areas, including areas where their homes were located); DANIEL A. FARBER ET AL., CASES AND MATERIALS ON CONSTITUTIONAL LAW 72 n.k (3d ed. 2003) (*Korematsu* was the last Supreme Court decision to permit overt discrimination against a racial or ethnic minority.); *see also* Adarand Constructors, Inc. v. Pena, 515 U.S. 200, 275 (1995) (Ginsburg, J., dissenting) ("A *Korematsu*-type classification, as I read the opinions in this case, will never again survive scrutiny: Such a classification, history and precedent instruct, properly ranks as prohibited.").

204. *Johnson*, 543 U.S. at 515 ("Prisons are dangerous places, and the special circumstances they present may justify racial classifications in some contexts. Such circumstances can be considered in applying strict scrutiny, which is designed to take relevant differences into account.").

205. *See id.*

206. The Court distinguishes the CDC policy where cellmates are not integrated from the cellmate policies in Texas and Oklahoma. *See supra* note 117 and accompanying text. Although the Court implies that the Oklahoma policy of individual assessment is theoretically more narrowly tailored than California's cellmate policy, the Court is careful to disclaim they have no knowledge of how the policy functions in reality. *Johnson*, 543 U.S. at 508 n.2.

housing policy is likely over-inclusive. If rival gang members are matched as cellmates because they share the same race, then a race-based housing policy is likely under-inclusive. By comparing the methodology of race-based housing policies with its actual results, courts can measure the extent to which a race-based housing policy is narrowly tailored.[207]

In *Johnson*, the United States Supreme Court extended strong constitutional protection under the Fourteenth Amendment to prisoners who face race-based policies.[208] The Court reaffirmed its commitment to rigidly scrutinizing racial classification in the unlikely context of the prison system,[209] where inmates traditionally have limited constitutional protection,[210] and prison officials traditionally enjoy deferential review.[211] To follow the *Johnson* standard of review, courts must carefully scrutinize the methodology by which prison administrators racially assign cellmates and the data for indications that an assignment policy is over- or under-inclusive.

DENIS ARCHIBALD FALLON

207. *See generally* Carter, *supra* note 51, at 433-34 (discussing the differences between laws designed to benefit and burden minorities).

208. *Johnson*, 543 U.S. at 515.

209. *See id.* at 524 (Thomas, J., dissenting) ("The Constitution has always demanded less within the prison walls.").

210. *Id.*

211. *Id.*; Washington v. Harper, 494 U.S. 210, 224 (1990) (finding "the [relaxed] standard of review we adopted in *Turner* applies to all circumstances in which the needs of prison administration implicate constitutional rights").

FROM STATUS TO CONTRACT[1]: EVOLVING PARADIGMS FOR REGULATING CONSUMER CREDIT

RASHMI DYAL-CHAND[2]

INTRODUCTION

In the last four decades something radical has happened in the United States consumer economy: The ordinary, middle class homeowner has gained a new means for borrowing thousands of dollars. While consumers[3] have long had the option of using their homes as collateral for loans,[4] they now can use credit cards to borrow large sums without any security for those loans.[5]

---

1. According to Sir Henry Sumner Maine (1822-1888), "the movement of the progressive societies has hitherto been a movement *from Status to Contract*." HENRY SUMNER MAINE, ANCIENT LAW: ITS CONNECTION WITH THE EARLY HISTORY OF SOCIETY AND ITS RELATION TO MODERN IDEAS 165 (10th ed. 1963) (1861) (alteration in original). This article suggests that such a movement does not necessarily signify progress in the area of consumer borrowing.

2. Associate Professor of Law, Northeastern University School of Law. I am grateful to Judith Olans Brown, Peter Enrich, Benjamin Ericson, Ingrid Michelsen Hillinger, Grant Nelson, Daniel Schaffer, and Joseph William Singer for their invaluable comments to prior drafts, and to James Hackney, Cathy Lesser Mansfield, David Phillips, and Elizabeth Warren for stimulating conversations about this project. I received terrific research assistance from Matthew Greene, Elizabeth Leduc, Michelle Moor, and Aimee Owen, unparalleled administrative support from Jan McNew, and unending research support from the staff at the Northeastern University Law Library.

3. This article uses the term "consumer" to distinguish between individuals and businesses. I focus exclusively on individual borrowing and developments in consumer credit in this country.

4. While other means of consumer borrowing, such as bank signature loans, installment credit, and small unsecured loans from consumer finance companies, did exist prior to the invention of the credit card, see ROBERT D. MANNING, CREDIT CARD NATION 3-4 (Basic Books 2000); NATIONAL CONSUMER LAW CENTER, STOP PREDATORY LENDING 4-6 (2002) (hereinafter STOP PREDATORY LENDING), mortgages were by far the most ubiquitous and traditional means of consumer borrowing. *See* RALPH J. ROHNER & FRED H. MILLER, TRUTH IN LENDING 245 (Robert A. Cook et al., eds., 2000).

5. My claim that credit card borrowing is approximately forty years old is based on the invention in the late 1950s of BankAmericard, Visa, and MasterCard credit cards, which were the first general purpose credit cards to allow borrowers to carry balances from month to month. DAVID S. EVANS & RICHARD SCHMALENSEE, PAYING WITH PLASTIC: THE DIGITAL REVOLUTION IN BUYING AND BORROWING 10-11 (The MIT Press 1999); MANNING, *supra* note 4, at 84-86; TERESA A. SULLIVAN ET AL., THE FRAGILE MIDDLE CLASS 109 (Yale University Press 2000); Oren Bar-Gill, *Seduction by Plastic*, 98 NW. U. L. REV. 1373, 1381 (2004). Arguably, however, credit card borrowing began several decades earlier with the introduction of merchant-specific credit cards, such as the card Sears Roebuck & Company offered. MANNING, *supra* note 4, at

This article explains the legal shift that accompanied this new type of loan transaction and explores changes in the credit relationship. While careful and extensive legal scholarship draws attention to the dramatic changes the credit card has made to the landscape of consumer credit, the focus has been on explaining the lure of credit cards to consumers[6] and the key legal moments that promoted the credit card.[7] The effect of credit cards on borrowers' legal rights has remained largely unexamined.[8]

This article seeks to fill that gap. Using mortgage lending as the historical paradigm, I claim that credit card lending creates a new legal paradigm of an individual borrower. Whereas mortgage law crafted non-waivable property rules to protect the ignorant[9] borrower from exploitation by the lender, credit

---

83-84; Bar-Gill, *supra*, at 1381.

6. *See, e.g.*, LLOYD KLEIN, IT'S IN THE CARDS: CONSUMER CREDIT AND THE AMERICAN EXPERIENCE (Praeger Publishers 1999); LEWIS MANDELL, THE CREDIT CARD INDUSTRY: A HISTORY (Twayne Publishers 1990); JAMES MEDOFF & ANDREW HARLESS, THE INDEBTED SOCIETY: ANATOMY OF AN ONGOING DISASTER (Little, Brown and Co. 1996); Bar-Gill, *supra* note 5; Laurie A. Lucas, *Integrative Social Contracts Theory: Ethical Implications of Marketing Credit Cards to U.S. College Students*, 38 AM. BUS. L.J. 413 (2001); Todd J. Zywicki, *The Economics of Credit Cards*, 3 CHAP. L. REV. 79 (2000); Ronald J. Mann, *Global Credit Card Use and Debt: Policy Issues and Regulatory Responses* (Apr. 2005), http://papers.ssrn.com/abstract'509063. Relatedly, important contributions have been made regarding the impact of credit cards on American society. *See, e.g.*, MANNING, *supra* note 4; SULLIVAN ET AL., *supra* note 5.

7. *See, e.g.*, Christopher C. DeMuth, *The Case Against Credit Card Interest Rate Regulation*, 3 YALE J. ON REG. 201 (1986); Margaret Howard, *Shifting Risk and Fixing Blame: The Vexing Problem of Credit Card Obligations in Bankruptcy*, 75 AM. BANKR. L.J. 63 (2001); Christopher L. Peterson, *Truth, Understanding, and High-Cost Consumer Credit: The Historical Context of the Truth in Lending Act*, 55 FLA. L. REV. 807 (2003); Vincent D. Rougeau, *Rediscovering Usury: An Argument for Legal Controls on Credit Card Interest Rates*, 67 U. COLO. L. REV. 1 (1996); Elizabeth R. Schiltz, *The Amazing, Elastic, Ever-Expanding Exportation Doctrine and Its Effect on Predatory Lending Regulation*, 88 MINN. L. REV. 518 (2004).

8. Treatises and similar summaries of law provide the most extensive treatment of credit card borrowers' rights. *See, e.g.*, MARK BUDNITZ & MARGOT SAUNDERS, NATIONAL CONSUMER LAW CENTER, CONSUMER BANKING AND PAYMENTS LAW (2d ed. 2002); BARKLEY CLARK & BARBARA CLARK, THE LAW OF BANK DEPOSITS, COLLECTIONS AND CREDIT CARDS (1999); NAT'L CONSUMER LAW CTR., CONSUMER CREDIT LAW MANUAL (1998) [hereinafter CONSUMER CREDIT LAW MANUAL]; ROHNER & MILLER, *supra* note 4; IRVING J. SLOAN, THE LAW AND LEGISLATION OF CREDIT CARDS: USE AND MISUSE (1987).

9. The term "ignorant" directly reflects legal treatment of residential mortgagors and is not intended to be pejorative. *See, e.g.*, Earls v. Chase Bank of Tex., 59 P.3d 364, 368 (Mont. 2002) (finding that "the [statutory] acknowledgment of notice requirement was intended to protect less sophisticated borrowers" (citation omitted)); Westervelt v. Gateway Fin. Serv., 464 A.2d 1203, 1208 (N.J. Super. Ct. Ch. Div. 1983) (stating that the purpose of New Jersey's Secondary Mortgage Loan Act is "to protect the ignorant, the choiceless and the often overreached"); S. REP. NO. 103-169, at 22 (1993) (stating that "[e]vidence before the Committee indicates that some high-rate lenders are using non-purchase money mortgages to take

card law relies on largely waivable contract rules to afford the enlightened borrower greater access to credit. Mortgage law traditionally protected ignorant borrowers' ownership rights in their collateral. Credit card law regulates the disclosures made to enlightened borrowers to ensure that they have enough information to make good bargains. This switch in borrower paradigms elicits drastically different results and less favorable treatment of the typical, middle class[10] consumer borrower in courts, Congress, and the market.

Part I explores the development of the ignorant borrower paradigm in mortgage law and the enlightened borrower paradigm in credit card law. The two paradigms are historical products of the social and legal policies of their days. Their coexistence, however, is not a logical response to consumer behavior in the market for credit. For the consumer, the choice between borrowing regimes is much more subtle, and the losses can be equally devastating. Consequently, the treatment of the borrower as ignorant in one instance and enlightened in the other is far less rational.

Having established the existence of the two paradigms, Part II explains the consequences of this paradigm shift. Part II.A. discusses the effect of the two paradigms on borrowers' rights, remedies, and days in court. Briefly, judicial use of mortgage law's ignorant borrower paradigm facilitates a more contextualized review of a credit transaction and creates opportunity for specific and meaningful remedies. The contract language in credit card cases, however, constrains judicial review; as long as fraud, unconscionability, or similarly egregious behavior does not occur in the formation of the contract, the contract provisions limit the borrowers' rights. Part II.B. suggests that what the mortgage market deems aberrant and predatory behavior the credit card market considers mainstream and acceptable. The enlightened borrower paradigm contributes fundamentally to this redefinition of lending norms. Part II.C., however, shows that the law responds to this important shift in the credit market with a continuing reliance on disclosure to remedy abuses in the credit card market.

Part III considers the future. The current legal environment in which the two borrower paradigms coexist is unstable, and disclosure appears to be the favored form of regulation. In continuing to respond to credit card borrowing, lawmakers should focus more attention on the role of this new form of credit in accomplishing important social goals.

---

advantage of unsophisticated, low income homeowners" in the legislative history of the Home Ownership and Equity Protection Act of 1994). For historical examples in which courts expressed these sentiments, see Vernon v. Bethell, 28 Eng. Rep. 838, 839 (Ch. 1762); Conway's Ex'rs & Devisees v. Alexander, 11 U.S. (7 Cranch) 218, 237 (1812); Youle v. Richards, 1 N.J. Eq. 534, 538 (N.J. Ch. 1832); Samuel v. Jarrah Timber & Wood-Paving Corp., 73 L. J. R. 526, 527-28 (Ch. 1904).

10. This article focuses on the effect of the paradigm shift on middle and lower middle class individuals rather than on the rich or poor.

I. THE EMERGENCE OF TWO PARADIGMS

Statistics on individual debt in this country abound. Sixty-nine percent of Americans own their own homes,[11] the vast majority financed with mortgages.[12] In addition, home equity borrowing[13] is a growing industry, totaling $415 billion by the middle of 2003.[14] Meanwhile, as of 2000, approximately seventy-seven percent of American households had at least one credit card.[15] Of the seventy-five percent of these individuals who carry balances from month to month,[16] the average credit card debt is a shocking $12,500.[17] These numbers suggest that many Americans carry significant amounts of both mortgage and credit card debt. As described by Professors Sullivan, Warren and Westbrook, "From 1980 to 1994, total household debt had increased from 65 percent to 81 percent of total income. In short, real

---

11. U.S. Census Bureau, Housing Vancancies and Homeownership, http://www.census.gov/hhes/www/housing/hvs/annual04/ann04t12.html. Moreover, those below the poverty line own approximately 8% of all homes in the United States. U.S. Census Bureau, American Housing Survey for the United States: 2003, http://www.census.gov/hhes/www/housing/ahs/ahs03/tab212.htm. The National Consumer Law Center recently found that over 44% of low-income Americans own their own homes. STOP PREDATORY LENDING, *supra* note 4, at 17.

12. According to 2003 Census figures, approximately 65% of homeowners had mortgages of some sort on their homes. U.S. Census Bureau, American Housing Survey for the United States: 2003, http://www.census.gov/hhes/www/housing/ahs/ahs03/tab315.htm [hereinafter U.S. Census Bureau]. Moreover, Professors Sullivan, Warren, and Westbrook state, on the basis of 1991 Census figures, that "[m]ore than 90 percent of homebuyers need a mortgage to purchase a home." SULLIVAN ET AL., *supra* note 5, at 206.

13. A home equity loan is "secured by a homeowner's residence other than loans used solely to purchase or construct the residence, to refinance a purchase money loan, or to make home improvements." Julia Patterson Forrester, *Mortgaging the American Dream: A Critical Evaluation of the Federal Government's Promotion of Home Equity Financing*, 69 TUL. L. REV. 373, 377 (1994). Consumers use such loans to finance everything from school tuition to groceries or luxury items.

14. David Myron, *Home Equity Debt Soars*, AMERICAN DEMOGRAPHICS (Nov. 1, 2004), available at http://www.findarticles.com/p/articles/mi_m4021/is_9_26/ai_n6344044 (relying on statistics maintained by the Federal Deposit Insurance Corporation). Home equity loan growth outpaced even credit card borrowing. Id. 2003 Census data show that home equity lump-sum mortgages, lines of credit, or both encumber more than 15% of homeownership in the United States. U.S. Census Bureau, *supra* note 12.

15. MANNING, *supra* note 4, at 317 n.11. This figure refers to "bank" or all-purpose credit cards. *Id.* Manning also notes that the average number of cards per cardholder is ten, including bank, retail, phone, gasoline, and travel and entertainment cards. *Id.* at 6. *See also* Bar-Gill, *supra* note 5, at 1384 (stating on the basis of a 1995 study "that 80% of [American] households have at least one credit card").

16. SULLIVAN ET AL., *supra* note 5, at 110-11.

17. Bar-Gill, *supra* note 5, at 1385.

consumer debt has risen dramatically over a long period during which real incomes for many people have stayed the same or declined."[18]

But the law does not treat all consumer debt alike. This section compares the law that governs the two major regimes of consumer borrowing, mortgages and credit cards.[19] The law has played an integral role in promoting these two forms of debt, but they are not products of the same law or promoted by similar policies. Consumer advocates and scholars alike have noted a deregulatory impulse in both arenas, which many argue results in significant exploitation of borrowers. Yet, the law has managed the difficult balance between access and exploitation very differently in these two arenas. These differences, despite being significantly consequential to the average consumer, often are not recognized.

A. Mortgages: The Ignorant Borrower

A cursory description of the history of mortgage law is that it consists of almost four hundred years of regulation designed to protect borrowers from loss of their collateral and approximately twenty-five years of deregulation designed to empower borrowers to greater participation in the market for credit.[20] The

---

18. SULLIVAN ET AL., *supra* note 5, at 18.

19. The purpose of this section is not to provide a comprehensive legal history of either credit cards or mortgages. Others have served that purpose well. For a legal history of mortgages, see, e.g, GERALD KORNGOLD & PAUL GOLDSTEIN, REAL ESTATE TRANSACTIONS, ch. 5 (4th ed. 2002); GRANT S. NELSON & DALE A. WHITMAN, REAL ESTATE FINANCE LAW, §§ 1.1-1.7 (4th ed. 2002) [hereinafter REAL ESTATE FINANCE LAW]; GEORGE E. OSBORNE, HANDBOOK ON THE LAW OF MORTGAGES, §§ 1-16 (2d ed. 1970); BERNARD RUDDEN & HYWEL MOSELEY, AN OUTLINE OF THE LAW OF MORTGAGES 3-10 (1967); Ann M. Burkhart, *Lenders and Land*, 64 MO. L. REV. 249 (1999); Morris G. Shanker, *Will Mortgage Law Survive? A Commentary and Critique on Mortgage Law's Birth, Long Life, and Current Proposals for Its Demise*, 54 CASE W. RES. L. REV. 69 (2003); David Sugarman & Ronnie Warrington, *Land Law, Citizenship, and the Invention of "Englishness": The Strange World of the Equity of Redemption, in* EARLY MODERN CONCEPTIONS OF PROPERTY, 111-143 (John Brewer & Susan Staves eds., 1995) (focusing solely on history of English mortgage law, the model for American mortgage law). For a legal history of credit cards, see, e.g, MANNING, *supra* note 4; SULLIVAN ET AL., *supra* note 5, at 109-11; Bar-Gill, *supra* note 5, at 1380-83.

20. While not a result of deregulation, the shift towards greater market participation arguably started shortly after the Great Depression, which caused an unprecedented number of foreclosures and incidents of homelessness. In response, the government created the Federal Housing Administration (FHA) in 1934, followed by the Federal National Mortgage Association (commonly referred to as Fannie Mae) in 1938. These developments helped support the secondary mortgage market and made mortgages available to more borrowers. Burkhart, *supra* note 19, at 272-74; Cathy Lesser Mansfield, *The Road to Subprime "HEL" Was Paved with Good Congressional Intentions: Usury Deregulation and the Subprime Home Equity Market*, 51 S.C. L. REV. 473, 477 (2000). Efforts to provide homeownership opportunities to World War II veterans also contributed to increased access. In particular, FHA and Veterans' Administration-insured loans made mortgage credit easier to attain. *See* Mansfield, *supra*, at 480-484; U.S. Department of Housing and Urban Development, The

regulatory phase appears to have been a process of accretion,[21] in which courts, and later legislatures, sought to limit lenders' various forms of exploitation. These lenders typically had more bargaining power, resources, and sophistication than their borrowers. The law during this period evidences a clear theme: State courts and legislatures consistently overrode mortgage contract terms to help borrowers retain their collateral. Despite the recent deregulation of the consumer mortgage market, some pro-borrower sensitivity remains today in the law of mortgage lending.[22]

The earliest and perhaps greatest achievement of the regulatory phase was the creation of the equity of redemption. Before the equity of redemption existed, a mortgage lender typically required the borrower to transfer to the lender the borrower's fee title to the collateral at the time of contract formation.[23] Common law courts strictly enforced this transfer, even when the loan terms were extremely onerous, unless the borrower repaid the loan according to the terms of the contract.[24] For example, where the loan agreement required a borrower to make a lump sum payment on a given date and the borrower violated the agreement merely by repaying the loan on the day after, the default resulted in his[25] loss of the collateral.[26]

Federal Housing Administration, http://www.hud.gov/offices/hsg/fhahistory.cfm.

21. The regulatory period was initiated by English courts of equity and ultimately supplemented by state legislatures in the United States. *See generally* OSBORNE, *supra* note 19, at 12-25 (providing historical background).

22. Importantly, as predatory lending practices demonstrate, this sensitivity is not necessarily reflected in the practice of mortgage lending, as opposed to the law. *See infra* Part II.B. As discussed in greater detail in Part III.A., the abuses in the market for consumer mortgages argue in favor of more stringent pro-borrower regulations.

23. Under the common law rule, the lender received all of the incidents of fee simple title, even though this was "antagonistic to the sole purpose of the conveyance, namely, security for the performance of an act by the grantor. Thus . . . creditors of the grantee could seize the property itself; . . . and [they] had the right to immediate possession." OSBORNE, *supra* note 19, at 9-10. Title theory states still write mortgages to effectuate an absolute conveyance, but the equity of redemption and other reforms render this a nominal conveyance, regardless of whether the state follows the title theory, lien theory, or a hybrid theory of mortgage law. *Id.* at 24-25. The basic distinction between title and lien theory states is that the former view title as passing to the mortgagee upon creation of a mortgage in accordance with English common law, while the latter view title as remaining with the mortgagor until foreclosure. *Id. See also* REAL ESTATE FINANCE LAW, *supra* note 19, at 10.

24. REAL ESTATE FINANCE LAW, *supra* note 19, at 7; OSBORNE, *supra* note 19, at 12-13; RUDDEN & MOSELEY, *supra* note 19, at 5; *see also* Shanker, *supra* note 19, at 73 (noting that prior to the advent of the equity of redemption "freedom of contract was paramount" and the "courts enforced [the agreed upon bargain] literally and precisely").

25. I use the male pronoun throughout this section in order to reflect the historical reality of mortgage borrowing.

26. "Law day" refers to the date of payment and this common law rule was absolute: Failure to pay on law day led to total forfeiture of the property. REAL ESTATE FINANCE LAW, *supra* note 19, at 7.

Although law courts strictly enforced such terms, courts of equity provided relief to borrowers in circumstances such as those described above, and, over time, in an increasingly broad range of circumstances. Where the borrower failed to pay on time, for reasons ranging initially from mistake to *force majeure*,[27] and ultimately to tardiness without justification,[28] the borrower's eventual late payment in full "redeemed" his right to the collateral. The equity of redemption exists in every state in the United States.[29] It overrides any contractual language attempting to limit or negate it, such as a waiver[30] or time limit.[31] The equity of redemption is a bright-line limitation on the lender's ability to use the collateral for any purpose other than to recoup the amount lent in the event that the borrower is conclusively unable to pay.

The counterbalance to the borrower's equity of redemption was, and still is, the lender's right of foreclosure.[32] But, over time, laws heavily regulated this lenders' remedy in order to protect the borrower.[33] Today in all but two states,[34] foreclosure requires the lender to go through a public sale to obtain[35]

---

27. OSBORNE, *supra* note 19, at 13.

28. *Id.*

29. CAROLYN L. CARTER ET AL., NATIONAL CONSUMER LAW CENTER, REPOSSESSIONS AND FORECLOSURES 511, 514 (5th ed. 2002) [hereinafter, REPOSSESSIONS AND FORECLOSURES]; Marshall E. Tracht, *Renegotiation and Secured Credit: Explaining the Equity of Redemption*, 52 VAND. L. REV. 599, 610 (1999). In addition, courts and legislatures carefully guard against "clogging" of the equity of redemption. *See* REAL ESTATE FINANCE LAW, *supra* note 19, at 34; Shanker, *supra* note 19, at 75.

30. *See, e.g.*, Cowley v. Shields, 60 So. 267, 269 (Ala. 1912) ("There is no merit in the contention that the warranty in the mortgage would preclude the mortgagor or his assignee from redeeming the property; for, if there is such a warranty, it could not operate to cut off the redemption right. Such a warranty or agreement is against public policy."); *see also* REAL ESTATE FINANCE LAW, *supra* note 19, at 35 (explaining the clogging doctrine and waiver); Shanker, *supra* note 19, at 75 (stating that "a mortgagor's equity of redemption could not be made irredeemable").

31. *See, e.g.*, Frazer v. Couthy Land Co., 149 A. 428, 429 (Del. Ch. 1929) (holding that despite a "clause in the [mortgage] agreement . . . , whereby the complainant agreed to the time limit of three years as the outside boundary of his right to take the land back . . . a right to redeem continues at all times until it is foreclosed"); *see also* REAL ESTATE FINANCE LAW, *supra* note 19, at 35.

32. *See* REAL ESTATE FINANCE LAW, *supra* note 19, at 7 (stating that the equity of redemption and right to foreclosure developed concurrently).

33. Eventually, some of these regulations were removed. For example, although foreclosure prototypically required court action, non-judicial foreclosure is now allowed in 60% of jurisdictions. REAL ESTATE FINANCE LAW, *supra* note 19, at 512.

34. Most states no longer allow strict foreclosure, a process by which the mortgagee receives title upon the mortgagor's default without having to bid at a sale, though Connecticut and Vermont continue to use this procedure, and other states allow it in certain circumstances. *See* CONN. GEN. STAT. ANN. § 8-256 (West 2001); 735 ILL. COMP. STAT. ANN. 5/15-1403 (West 2003) (allowing strict foreclosure only when three conditions are met); VT. STAT. ANN. tit. 9A, § 9-620 (2001). For general discussions of strict foreclosure, see KORNGOLD & GOLDSTEIN, *supra* note 19, at 458; REAL ESTATE FINANCE LAW, *supra* note 19, at 554-57.

fee title to the collateral.[36] Moreover, state statutes extensively control the legal steps required to accomplish such a sale. In the forty percent of states that allow only judicial foreclosure, such regulations typically include requirements of adequate notice,[37] a hearing,[38] regulation of the method of sale, and a process for determining rights to any surplus.[39] But even those states allowing non-judicial foreclosure, most prevalently the power of sale, regulate notice and other aspects of the foreclosure process.[40] Like the equity of redemption, the loan agreement cannot alter these statutory limits on foreclosure. For example, the loan agreement cannot expand the lender's rights with remedies, such as strict foreclosure or a deed in lieu of foreclosure,[41] and cannot limit the borrower's rights in foreclosure, such as by curtailing the notice to the borrower.[42]

---

35. In title theory jurisdictions, the lender technically retains fee title to the collateral, giving the lender the right of possession, but most title states treat it simply as a security interest. *See* REAL ESTATE FINANCE LAW, *supra* note 19, at 130-33. Regardless of the type of jurisdiction, the practical consequences to the borrower and lender are largely the same. *See supra* note 23.

36. In other words, the remedy of foreclosure is not aimed at giving the lender title to the collateral, but rather at giving the lender proceeds from the sale of the collateral. Burkhart, *supra* note 19, at 269 (noting "even the title theory states recognize that the lender's paramount right is to repayment of the debt"); *see also* Shanker, *supra* note 19, at 76 (explaining the evolution from strict foreclosure to foreclosure by sale).

37. According to Osborne, "The purpose of notice is to inform the public as to the date, place, nature and condition of the property to be sold and terms of the sale. Questions as to the sufficiency and definiteness of notice to accomplish these purposes sometimes are of importance." OSBORNE, *supra* note 19, at 683.

38. REAL ESTATE FINANCE LAW, *supra* note 19, at 492.

39. *Id.* Importantly, however, after foreclosure, it is difficult to overturn a foreclosure sale on grounds such as the inadequacy of the sales price. *See id.* at 589.

40. *Id.* at 513; OSBORNE, *supra* note 19, at 733-36.

41. Both of these remedies are also attempted clogs on the equity of redemption, which mortgage law abhors. *See, e.g.*, First Illinois Nat'l Bank v. Hans, 493 N.E.2d 1171, 1174-75 (Ill. App. Ct. 1986) (refusing to enforce a clause in the mortgage "requir[ing] the mortgagor to execute a deed upon the occurrence of a default" (emphasis omitted)); *see also* Steven Wechsler, *Through the Looking Glass: Foreclosure by Sale As* De Facto *Strict Foreclosure - An Empirical Study of Mortgage Foreclosure and Subsequent Resale*, 70 CORNELL L. REV. 850, 851, 858 (1985) (discussing the relationship between strict foreclosure and the equity of redemption, but arguing, on the basis of empirical evidence, that strict foreclosure should be permitted so long as additional mortgagor protections are in place). While deed in lieu clauses in the original loan agreement are not enforceable, the law does not necessarily prohibit them in subsequent transactions. Accordingly, the parties may later agree to a deed in lieu of foreclosure, though the courts carefully scrutinize such agreements for signs of fraud or oppression. *See* REAL ESTATE FINANCE LAW, *supra* note 19, at 43.

42. *See, e.g.*, Fleisher Co. v. Grice, 226 A.2d 153, 155 (Md. 1967) (stating that parties are free to contract beyond the base amount of notice required by the statute but not below). Courts regularly approve of foreclosure proceedings where a reasonable attempt was made to follow the statute, even if strict compliance did not occur. *See* REPOSSESSIONS AND FORECLOSURES, *supra*

Over half of the state legislatures have also added a statutory right of redemption that allows the borrower[43] to redeem the property for a certain period of time *after* the foreclosure.[44] In addition, by the 1970's, many state legislatures had passed usury laws capping interest rates that could be charged on residential mortgage loans.[45]

Courts have long supplemented these borrower-protective remedies with a range of interventions focusing on contract formation and interpretation, which effectively trump the contract language. A textbook example is that of a contract for an outright conveyance of land that the court determines to be a mortgage transaction on the basis of parol evidence.[46] Often, this rewriting of the contract provides the borrower with the benefit of the equity of redemption,[47] sometimes it limits the interest rates charged,[48] and sometimes it imposes foreclosure rules on the lender.[49]

As these legal rules illustrate, one characteristic of the residential mortgagor is that he is a legal entity worth investing in, including by means of favorable legal rules. Another important characteristic is that he is a legal entity

---

note 29, at 515. In some states, there are also statutory limits on deficiency judgments that can be obtained by lenders whose loans are not fully satisfied by proceeds from the sale. For a history and description of anti-deficiency legislation, see REAL ESTATE FINANCE LAW, *supra* note 19, at 658-88. Here, too, even where the mortgage loan agreement provides for full recourse, the effect of some of these statutes is to transform these loans into non-recourse loans.

43. Junior lienors are also often permitted to redeem. REAL ESTATE FINANCE LAW, *supra* note 19, at 469. James B. Hughes, Jr., *Taking Personal Responsibility:A Different View of Mortgage Anti-Deficiency and Redemption Statutes*, 39 ARIZ. L. REV. 117, 130 n.81 (1997).

44. KORNGOLD & GOLDSTEIN, *supra* note 19, at 492-93; *see* REAL ESTATE FINANCE LAW, *supra* note 19, at 689-706. This right is in addition to the equity of redemption.

45. *See* Schiltz, *supra* note 7, at 526. But these usury laws were later largely preempted, as discussed below. *See infra* notes 54-60 and accompanying text.

46. *See* Part II.A.i. Justice Field's majority opinion in *Pierce v. Robinson* is a typical reaction to parol evidence in the mortgage context:

> I consider parol evidence admissible in equity, to show that a deed absolute upon its face was intended as a mortgage, and that the restriction of the evidence to cases of fraud, accident, or mistake, in the creation of the instrument, is unsound in principle and unsupported by authority. The entire doctrine of equity, in respect to mortgages, has its origin in considerations independent of the terms in which the instruments are drawn.

13 Cal. 116, 125 (1859); *see also* KORNGOLD & GOLDSTEIN, *supra* note 19, at 356 (stating that "[c]ourts generally favor debtors in their willingness to pierce the facade of the absolute deed and find an equitable mortgage"); Shanker, *supra* note 19, at 74-75 (stating that the contract could not provide the mortgagee any interest in the mortgagor's property once the loan was repaid in full).

47. *See, e.g.*, Campbell v. Dearborn, 109 Mass. 130 (1872); *see also* REAL ESTATE FINANCE LAW, *supra* note 19, at 58 (stating that the right to redeem is a legal consequence of the mortgage's existence); OSBORNE, *supra* note 19, at 116 (stating that the right to equity of redemption "flow[s] from the existence of the mortgage relation").

48. *See, e.g.*, McElroy v. Grisham, 810 S.W.2d 933, 938 (Ark. 1991).

49. *See, e.g.*, Shadman v. O'Brien, 180 N.E. 532, 533 (Mass. 1932); Scheibe v. Kennedy, 25 N.W. 646, 649 (Wis. 1885).

susceptible to exploitation because of the borrower's level of power and sophistication as compared to the lender's level. Until the late 20th century, the legal paradigm of the residential mortgagor was that of the ignorant borrower who was unable to protect his interests by means of private ordering and who needed certain non-waivable legal rights and remedies to fully participate in the market for credit, as was necessary for the greater social good.[50]

The interventions developed during the regulatory phase included and exceeded the typical contract-based interventions, such as fraud, unconscionability, and the Statute of Frauds,[51] effectively superimposing property-based rules on the mortgage loan agreement. For example, the borrower does not have to bargain for the equity of redemption or for limitations on deficiency judgments[52]; these are rights associated with the status of mortgagor.[53] Consequently, mortgage transactions, in contrast to the typical non-land-based contract, have historically been the subject of extensive judicial interventions and statutory protections.

While at least partially intact today, this paradigm has degraded over the last twenty-five years as the focus has shifted from exploitation of the borrower towards access to credit.[54] The almost unanimous justification given for this deregulation was the need in the late 1970's and early 1980's to rejuvenate mortgage borrowing, which had become stagnant along with the economy.[55] Federal preemption of state laws regulating residential mortgage lending and consumer borrowing was perhaps the single greatest instrument of deregulation.[56] The most visible state laws to be preempted have been those

---

50. The particular social goods at issue here are those often associated with private property law, such as its contribution to wealth formation, to the creation of a market in real estate, and to the enhancement of land's alienability. *See infra* Part III.B.

51. The Statute of Frauds has a property-specific component: All contracts regarding land must be in writing. KORNGOLD & GOLDSTEIN, *supra* note 19, at 85.

52. *See, e.g.*, Mooney v. Byrne, 57 N.E. 163, 165 (N.Y. 1900) ("The right to redeem is an essential part of a mortgage, read in by the law, if not inserted by the parties."); Scheibe v. Kennedy, 25 N.W. 646, 647 (Wis. 1885) (stating that the same principle applies to the requirement of foreclosure). A mortgagor can bargain for a higher level of protection than the law allows. For example, borrowers can bargain for a fully non-recourse loan; however, the right of redemption and the requirements relating to foreclosure are examples of mortgagor protections on the basis of their status as "ignorant" borrowers.

53. Part I.D. discusses the status-based protections as property rules because they focus on protection of the borrower's ownership of collateral and have many characteristics of typical property rules, as defined by Calabresi and Melamed. *See* Guido Calabresi & A. Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089 (1972). In contrast, a contract approach requires the mortgagor to bargain for the right of redemption and for limitations on the right to foreclose.

54. *See* William N. Eskridge, Jr., *One Hundred Years of Ineptitude: The Need for Mortgage Rules Consonant with the Economic and Psychological Dynamics of the Home Sale and Loan Transaction*, 70 VA. L. REV. 1083 (1984).

55. *See* Mansfield, *supra* note 20, at 521.

56. Some statutes explicitly preempted state laws, while courts interpreted others as

capping usurious interest rates. Section 85 of the National Bank Act of 1864[57] authorizes national banks to charge interest at the highest rate charged by any lender in the state in which those national banks are located. State usury laws largely set the standard for permissible interest rates until two events occurred in the late 1970's. In 1978, the Supreme Court held that a national bank can "export" the interest rate permitted in the state in which the bank is "located"[58] to all states in which the bank's borrowers are located.[59] Consequently, many national banks located themselves in states such as South Dakota,[60] which had liberal usury laws, or none at all.[61] In 1979, Congress enacted the Depository Institutions Deregulation and Monetary Control Act of 1980 (DIDMCA), which explicitly preempted state usury laws.[62] The Alternative Mortgage

---

having a preemptive effect. *See* Schiltz, *supra* note 7, at 540-69.

57. 12 U.S.C. § 85 (2000).

58. The term "located" has been heavily litigated. *See* Schiltz, *supra* note 7, at 546-60.

59. Marquette Nat'l Bank of Minneapolis v. First of Omaha Serv. Corp., 439 U.S. 299, 318 (1978). Although *Marquette* concerns credit card lending, the Court makes numerous references to the "on any loan" language of the National Bank Act. *Id.* at 308, 313. Various courts have applied the *Marquette* doctrine to mortgage loans as well. *See, e.g.*, Wachovia Bank v. Burke, 319 F. Supp. 2d 275, 277 (D. Conn. 2004); Morequity, Inc. v. Naeem, 118 F. Supp. 2d 885, 897 (N.D. Ill. 2000); Kimball v. Nat'l. Bank of N. Am., 468 F. Supp. 1069, 1071 (E.D.N.Y. 1979).

60. South Datkota's applicable statute states:

> Unless a maximum interest rate or charge is specifically established elsewhere in the code, there is no maximum interest rate or charge, or usury rate restriction between or among persons, corporations, limited liability companies, estates, fiduciaries, associations, or any other entities if they establish the interest rate or charge by written agreement.

S.D. Codified Laws § 54-3-1.1 (2005).

61. Schiltz notes that liberal or non-existent state usury laws "give[ ] states such as South Dakota and Delaware incentive to engage in a 'race to the bottom' of consumer credit regulatory schemes, in order to attract consumer lending operations to their states." Schiltz, *supra* note 7, at 619. The South Dakota statute and the *Marquette* decision only apply to national banks. While many mortgage lenders are national banks, many are not. *See* Mansfield, *supra* note 20, at 526-27. By contrast, the National Bank Act and the *Marquette* decision have a broader preemptive effect in credit card lending. However, other federally chartered banks, state chartered banks, and their subsidiaries are also entitled to preemption of some state laws. Given the benefits of preemption to lenders, many other banking entities are finding ways to fall within the cover of the exportation clause. *See* NATIONAL CONSUMER LAW CENTER, THE COST OF CREDIT 67-69, 73 (2d ed. 2002); Schiltz, *supra* note 7, at 569-617.

62. Federal law states:

> The provisions of the constitution of any State expressly limiting the rate or amount of interest, discount points, or other charges which may be charged, taken, received, or reserved by lenders and the provisions of any State law expressly limiting the rate or amount of interest, discount points, or other charges which may be charged, taken, received, or reserved shall not apply to any loan, mortgage, or advance which is insured under subchapter I or II of this chapter.

12 U.S.C. § 1735f-7(a) (2000). *See also* Grant S. Nelson & Dale A. Whitman, *Reforming Foreclosure: The Uniform Nonjudicial Foreclosure Act*, 53 DUKE L.J. 1399, 1412 (2004)

Transactions Parity Act of 1982 (AMTPA) was another source of preemption, specifically of state laws requiring fixed interest rates on residential mortgage loans.[63]

Concurrent with the deregulatory effects of preemption, changes in tax laws,[64] creation of the secondary market in residential mortgage loans,[65] and massive entry by mortgage brokers and nondepository lenders into the residential mortgage market[66] succeeded in ensuring very broad access to mortgage credit.

Deregulation is often blamed for enabling predatory lending which occurs on a massive scale in the mortgage market today.[67] If contemporaneousness is

---

[hereinafter *Reforming Foreclosure*] (discussing the history of federal statutes preempting state usury laws).

63. 12 U.S.C. §§ 3803-04 (2000). Some of the new mortgages permitted by the AMTPA have been challenged as clogging the equity of redemption. *See* REAL ESTATE FINANCE LAW, *supra* note 19, at 36-38. A third statute, the Garn-St. Germain Depository Institutions Act of 1982, 12 U.S.C. § 1701j-3 (2000), also preempts state mortgage law. That statute makes due-on-sale clauses more uniformly enforceable by lenders. *Reforming Foreclosure*, *supra* note 62, at 1413. Office of the Comptroller of the Currency (OCC) and Office of Thrift Supervision (OTS) rulings and letters have extended the reach of these statutes to more entities and to more substantive provisions of the loan agreement, such as penalties for late payment and closing costs. Schiltz, *supra* note 7, at 535, 561-64, 602. In *Smiley v. Citibank*, 517 U.S. 735, 747 (1996), the Supreme Court upheld the OCC's significant expansion of the term "interest" in the National Bank Act as reasonable, finding that the OCC has the authority to interpret the definition in the statute it is charged with enforcing.

64. According to Professor Mansfield, the Tax Reform Act of 1986 "disallowed the deductibility of consumer interest but permitted taxpayers to deduct interest paid on loans secured by the taxpayer's principal and one other residence." Mansfield, *supra* note 20, at 522 (internal citations omitted).

65. Mansfield, *supra* note 20, at 531-32. The rise of the secondary market was undoubtedly aided by the greater uniformity resulting from preemption. *See Reforming Foreclosure, supra* note 62, at 1411-13. The holder in due course doctrine also contributed to the protections afforded purchasers of new securities. *See* THE COST OF CREDIT, *supra* note 61, at 422-23; Schiltz, *supra* note 7, at 537-38;.

66. Mansfield notes that "[t]his steady growth of finance company participation in real estate lending is significant because these lenders are now the most active group of subprime, nonpurchase money lenders." Mansfield, *supra* note 20, at 526.

67. *See, e.g.*, THE COST OF CREDIT, *supra* note 61, at 60-66; Mansfield, *supra* note 20, at 528-31. In 1993, approximately 100,000 subprime home loans were originated, and by 2002 the number of subprime home loans originated had increased to a staggering 1.36 million. Association of Community Organizations for Reform Now [hereinafter the ACORN Report], *Separate and Unequal: Predatory Lending in America* 6 (2004), http://www.acorn.org/index.php?id'1994. *See Legislative Solutions to Abusive Mortgage Lending Practices: Joint Hearing Before the Subcommittee on Financial Institutions and Consumer Credit, Subcommittee on Housing and Community Opportunity*, 109th Cong. at 6-7 (May 24, 2005) (statement of Martin Eakes, C.E.O., Self-Help and the Center for Responsible Lending). One of the major forces driving the subprime market was the massive influx of private capital and investors during the 1990's. "Subprime mortgage-backed securities grew

any indication, the legal counterbalances to the deregulation were the federal Consumer Credit Protection Act of 1968 (CCPA)[68] and the Real Estate Settlement Procedures Act of 1974 (RESPA).[69] The CCPA regulates disclosure,[70] billing,[71] debt collection,[72] and credit reporting[73] practices and addresses lending discrimination,[74] providing significant protection to consumer borrowers in all contexts, not just mortgage borrowing. These two statutes also heralded the dawn of a markedly different approach to consumer protection than that developed by mortgage law. The regulatory focus was on the adequacy of the disclosure made to the borrower about the terms of the lending contract, rather than on directly protecting the borrower from loss of collateral.[75]

While the deregulatory phase cultivated the current "ownership society" by encouraging broader availability of credit, status-based protections are not things of the past in mortgage law. Status still has a powerful influence over lawmakers.[76] None of the new statutes override the equity or the statutory right of redemption,[77] leaving intact one of the greatest protections for borrowers.

---

from over $18 billion issued in 1995 to more than $134 billion issued in 2002." Margot Saunders & Alys Cohen, *Federal Regulation of Consumer Credit: The Cause or the Cure for Predatory Lending?* 9 (2003), http://jchs.harvard.edu/publications/finance/babc/babc_04-21.pdf.

68. 15 U.S.C. §§ 1601-1693r (2000).

69. 12 U.S.C. §§ 2601-17 (2000).

70. Title I of the CCPA, commonly referred to as the Truth in Lending Act, regulates disclosure of credit terms. 15 U.S.C. §§ 1601-1667f (2000).

71. Fair Credit Billing Act, 15 U.S.C. § 1666 (2000).

72. Fair Debt Collection Practices Act, 15 U.S.C. § 1692 (2000).

73. Fair Credit Reporting Act, 15 U.S.C. § 1681 (2000).

74. Equal Credit Opportunity Act, 15 U.S.C. § 1691 (2000).

75. Part III.A. discusses how the disclosure regime prevalent in credit card lending has begun to infect mortgage lending as well. *See infra* Part III.A.

76. Many protections are only available to the borrower who ends up in court before the lender begins a foreclosure action, particularly in states which permit non-judicial foreclosure. As predatory lending vividly demonstrates, much of the exploitation of consumer borrowers can only be remedied by a lawsuit, and many borrowers lack the resources to go to court. In other words, if a borrower does not contest the terms of an onerous mortgage arrangement in court, she may never have the benefit of some of the most potent protections available. This barrier is the basis of legitimate critique by consumer advocates.

77. Some scholars have argued that there have been inroads on the equity of redemption. *See, e.g.*, John C. Murray, *Clogging Revisited*, 33 REAL PROP. PROB. & TRUST J. 279 (1998); Shanker, *supra* note 19, at 78-80 (arguing that the Uniform Land Security Interest Act, ULSIA, § 211, 7A U.L.A. 403 (1985), and the American Law Institute's 1997 Restatement of Mortgage law "expressly authorize the mortgagee in the mortgage contract to validly bargain for an interest in the mortgagor's property, *in addition to* their rights to collect the full underlying debt" (emphasis added)); Lou J. Viverito, Comment, *The Shared Appreciation Mortgage: A Clog on the Equity of Redemption?*, 15 J. MARSHALL L. REV. 131 (1982). In addition, some worry that the Ney-Kanjorski bill, which seeks to limit predatory lending, is broad enough to preempt the equity of redemption and any other state consumer protection laws. *See, e.g.*,

Although significantly less regulated than judicial foreclosure, remedies such as the power of sale are still regulated by state statutes, which cannot be waived or limited by contract.[78] Regarding contract formation, state courts continue to void or rewrite onerous clauses and whole contracts to provide greater protections to the borrower.[79] Whether purposeful or not, the current law of mortgage lending enshrines a series of balances between broad access to credit and protection of the borrower's status as owner of the collateral,[80] between federal and state regulation, and between property-protective contract re-interpretation and opportunities for private ordering.[81]

Margot Saunders & Alys Cohen, *NCLC Response to* Dear Colleague Letter Supporting H.R. 1295, (Apr. 19, 2005), http://www.consumerlaw.org/initiatives/predatory_mortgage/content/responsetodearcolleague.pdf (critiquing the Ney-Kanjorski bill). Finally, some might argue that OCC regulations such as 12 C.F.R. § 34.4 could be broadly interpreted to preempt the equity of redemption and other state law protections for the borrower.

78. Notice requirements, for example, are less strict for power of sale foreclosures than for judicial foreclosures. REAL ESTATE FINANCE LAW, *supra* note 19, at 582. Courts have held, however, that failure to give proper notice for power of sale foreclosures may render the sale void. *See* REPOSSESSIONS AND FORECLOSURES, *supra* note 29, at 512.

79. *See, e.g.*, APAC-Mississippi, Inc. v. JHN, Inc., 818 So. 2d 1213, 1220-21 (Miss. Ct. App. 2002) (holding, on the basis of parol evidence, that an absolute deed was a mortgage); LaFond v. Rumler, 574 N.W.2d 40, 44-45 (Mich. Ct. App. 1997) (declaring void, as a restraint on alienation, an addendum to a land contract requiring buyer to share proceeds of resale with seller); Patterson v. Grace, 661 N.E.2d 580, 583-84 (Ind. Ct. App. 1996) (holding that the lower court did not err in using parol evidence to find an equitable mortgage); Basile v. Erhal Holding Corp., 538 N.Y.S.2d 831, 832 (N.Y. App. Div. 1989) (refusing to enforce a deed in lieu of foreclosure).

The problem of predatory lending in the mortgage market demonstrates that these legal measures, while important, are insufficient to adequately protect borrowers from lender exploitation. Even the CCPA, with its focus on adequate disclosures to borrowers, has diluted mortgage law's emphasis on the borrower's status as a basis for substantive protections. It is reasonable to wonder whether a status-based strategy would result in more effective protections for borrowers against predatory lenders. Professors Engel and McCoy's suitability proposal, which requires the creditor to assess the suitability of the loan for the borrower prior to making the loan, might be an example of such protection. *See* Kathleen C. Engel & Patricia A. McCoy, *A Tale of Three Markets: The Law and Economics of Predatory Lending*, 80 TEX. L. REV. 1255, 1317-1366 (2002) [hereinafter *A Tale of Three Markets*].

80. I join the many consumer advocates who argue that this series of balances is insufficient, given the rampant exploitation that occurs in the subprime mortgage market. Nonetheless, in contrast to the credit card market, the mainstream or prime mortgage market does, at some level, maintain the balances described above.

81. Karen M. Pence suggests in her empirical study that state borrower-protective foreclosure laws benefit defaulting borrowers to the detriment of non-defaulting borrowers. *See* Karen M. Pence, *Foreclosing on Opportunity: State Laws and Mortgage Credit* 28 (2003), http://www.federalreserve.gov/pubs/feds/2003/200316/200316pap.pdf. As federal law preempts some state borrower protections, they maintain broad access to credit, while state laws that are not preempted continue to protect defaulting borrowers. Although the combined effect of these

B. Credit Cards: The Enlightened Borrower

Such balance is not evident in credit card lending. Although exploitation of credit card borrowers is rampant and rising, the credit card market continues to experience a deregulatory trend. Laws that regulate credit card lending have always used a markedly different approach from traditional mortgage law. The prevailing theme in the law of credit card lending is not one of contract reinterpretation to aid borrowers.

Instead, credit card law regulates disclosure. This approach originated with the CCPA, the "all-purpose" consumer credit statute enacted in 1968.[82] The Truth In Lending Act (TILA), whose purpose is "to assure a meaningful disclosure of credit terms" and "to protect the consumer against inaccurate and unfair credit billing and credit card practices,"[83] is the most obvious manifestation of this new regime. TILA accomplishes this purpose by focusing on the disclosure of the terms and costs of the loan—namely of finance charges and annual percentage rates[84]--and by regulating advertising.[85] Generally, as long as lenders do not deceptively state the terms of the loan, the law will not curb the terms' substantive harshness.[86]

The amended TILA does have special protections for credit card borrowers.[87] Although these protections have been described as more

---

laws may not be sufficient to counteract predatory lending, it does suggest a balance between access and exploitation.

82. The CCPA applies to all forms of consumer credit and generally does not distinguish between credit cards and mortgages. It does regularly distinguish between closed-end credit, where a fixed amount typically is loaned up-front, and open-end credit, where the borrower borrows money in frequent, often irregular transactions, and amounts which are merged into a single outstanding amount of debt. *See* THE COST OF CREDIT, *supra* note 61, at 51. Unlike most mortgages which are closed-end lending arrangements, credit card lending typically is open-end financing, where the amount financed depends on the amount purchased per month and not repaid fully at the end of the month. *See* Bar-Gill, *supra* note 5, at 1395-96.

83. 15 U.S.C. § 1601 (2000). At least one commentator has described TILA as the "centerpiece" of the CCPA. Schiltz, *supra* note 7, at 533.

84. 15 U.S.C. § 1632 (2000).

85. 15 U.S.C. §§ 1662-65 (2000).

86. The only real limitation that TILA imposes on the substantive terms of the contract is the right of rescission, which is available only to consumers entering into non-purchase money mortgages on their homes and only for three days if the creditor has followed the statutory disclosure requirements; if the consumer does not receive the required disclosures, the right of rescission can last three years. *Id.* § 1635. *See generally* MARY DEE PRIDGEN, CONSUMER CREDIT AND THE LAW § 15:16-19, 20-21 (2004); ROHNER & MILLER, *supra* note 4, at 683; Deborah Freye et al., *Truth in Lending Developments*, 59 BUS. LAW 1125, 1130 (2004); Brian P. Richards, *The Right of Rescission: A Continuing Concern for Mortgage Lenders*, 111 BANKING L.J. 557 (1994). Thus, one of the few substantive provisions of TILA that has the effect of limiting contract terms, is available only to mortgage borrowers.

87. 15 U.S.C. §§ 1601-1615 (2000).

substantive,[88] they are actually little more than specific remedies for inadequate disclosure and for failure of contract formation at the most basic level, more akin to an unconscionability defense than to the substantive property rules of traditional mortgage law. A 1970 amendment, for example, prohibits unsolicited issuance of credit cards, an attempt to ensure that a meeting of the minds occurred in the contract formation process and to protect consumers from third party fraud, which the creditor is in a better position to prevent.[89] A second amendment limits consumers' liability for unauthorized use of their credit cards.[90] While this provision overrides any contract language to the contrary, it, too, is a protection against the most fraudulent type of activity, which, although not perpetrated by the creditor, is more easily prevented by the creditor than the borrower.[91] These protections are significantly less borrower-protective than the limitations on creditors' remedies and contract reinterpretation that have been the norm in mortgage law.

TILA's disclosure regime assumes it deals with a markedly different type of borrower than mortgage law's borrower. Most importantly, it assumes that once a borrower is properly informed of the credit terms proposed by the lender, such as the actual interest rate charged, the borrower will have enough sophistication and bargaining power to negotiate loan terms that are reasonably favorable. Also implicit in the non-mortgage-specific provisions of this statute is an assumption that the borrower is not risking very much because there is no home or other collateral at stake. In contrast to the traditional mortgage borrower, the TILA borrower needs very few rights that are non-waivable by contract. It is only when the lender is in a position to avoid the risk that the statute overrides traditional contract rights and remedies in favor of the borrower. In other words, in contrast to the ignorant borrower of traditional mortgage law, the prototypical borrower assumed by the TILA regime is enlightened.

While not explicitly focused on disclosure, the other pieces of the CCPA assume the same borrower paradigm as TILA. The Fair Credit Reporting Act (FCRA), for example, administers a system requiring credit bureaus to maintain and disseminate to certain types of creditors[92] accurate credit reports on individuals, which the credit bureaus must also disclose to those individuals upon request.[93] The FCRA does not regulate the ways in which a lender can

88. ROHNER & MILLER, *supra* note 4, at 685-87.

89. 15 U.S.C. § 1642 (2000). For an analysis of this amendment, see ROHNER & MILLER, *supra* note 4, at 688-93.

90. 15 U.S.C. § 1643 (2000).

91. Other protections for credit card borrowers include the requirement that credit card companies post payments on the day received under the TILA and greater rights in the billing context under the Fair Credit Billing Act. For analyses of these protections, see CONSUMER CREDIT LAW MANUAL, *supra* note 8, at § 15.

92. FCRA also applies to certain other entities, such as employers. 15 U.S.C. § 1681(b) (2000).

93. 15 U.S.C. § 1681g (2000).

use the credit report.[94] Meanwhile, the Fair Debt Collection Practices Act operates at the margin of private ordering by regulating collection practices, usually in the nature of tortious, unfair, or harassing behavior, that the borrower and lender may not have negotiated by contract.[95] While arguably more substantive than TILA, these statutes leave much of the loan contract fully intact, including onerous penalty clauses.[96]

Three other factors increase the dominance of the disclosure regime established by TILA in the context of credit card borrowing. Like most of TILA, these regulations do not isolate credit cards for special treatment; however, consumers feel their impact most widely in the context of credit card borrowing.[97] Importantly, none of these developments were driven by attention to the needs of, and risks borne by, credit card borrowers. For example, nowhere does the law appear to consider whether onerous penalty clauses limit

---

94. This is in contrast to FCRA's arguably more substantive regulation of an employer's use of credit reports, which includes requirements to disclose to the employee or potential employee her statutory rights and to obtain the employee's release signature before the employer can acquire the credit report. 15 U.S.C. § 1681(b) (2000). More broadly, the statute restricts the type of information that can appear on a consumer's credit report, such as certain medical information, and requires that those providing the information to creditors maintain accuracy and update a consumer's report when inaccuracies are known. 15 U.S.C. § 1681c (2000). The latter requirements, while not limiting lenders' behavior, could also be seen as substantive restrictions on lending.

95. 15 U.S.C. § 1692 (2000). The same could be said about the Equal Credit Opportunity Act, which regulates against discrimination. 15 U.S.C. §§ 1691-1691f (2000). As I define substantive intervention, the Fair Credit Billing Act is the most interventionist, and thus the most similar to traditional mortgage law, in that it overrides contract language concerning procedures for accurate billing. 15 U.S.C. § 1666 (2000). For example, § 1666b regulates the circumstances under which finance charges can be imposed on open-end credit accounts, regardless of the terms of the contract between the borrower and creditor. Similarly, § 1666 sets forth the procedure for correcting billing errors, regardless of whether private bargaining allows the creditor to do less. *See also* 12 C.F.R. § 226.

96. I am not claiming here that the CCPA is not an important consumer protection statute or that it is ineffective. Rather, my claim is that the CCPA implicitly assumes a different kind of borrower than that protected by mortgage law.

97. The FTC Credit Practices Rules, 16 C.F.R. §§ 444.1 to 444.5, provide some protections that arguably override private bargaining. For example, § 444.2 prohibits lenders from requiring consumer borrowers to sign a confession of judgment, a waiver of limitation of exemption from attachment such as a homestead exemption, an assignment of wages, or convey a "nonpossessory security interest in household goods other than a purchase money security interest." 16 C.F.R. § 444.2(a)(4) (2000). Section 444.4 limits late charges "on a payment, which payment is otherwise a full payment for the applicable period and is paid on its due date or within an applicable grace period, when the only delinquency is attributable to late fee(s) or delinquency charge(s) assessed on earlier installment(s)." 16 C.F.R. § 444.4(a) (2000). While these protections are helpful, they minimally protect against the most egregious and unconscionable behavior, which would likely be unfair trade practices under state statutes or common law.

credit card borrowing, whether they result in too much exploitation, or how they generally affect borrower behavior.

The first contributing factor is the kind of preemption that has been taking place in the mortgage market for the past three decades. While the DIDMCA's primary impact is on mortgages,[98] and the AMTPA applies only to mortgage loans,[99] the preemptive effect of § 85 of the National Bank Act, as interpreted by the Supreme Court in *Marquette*,[100] applies equally to credit card borrowing. Indeed, the loan transaction in *Marquette* was an open-end credit card transaction.[101] The second factor intensifies the effect of § 85 on the rights of credit card borrowers. Major credit card issuers are national depository institutions or their subsidiaries[102] and have benefited from significant deregulation of their practices in the past few decades.[103] To counterbalance the regulatory impositions on national banks, the Supreme Court, Congress, and the agencies responsible for regulating national banks have bestowed upon them rights equal to state banks' privileges in important substantive areas.[104]

---

98. 12 U.S.C. § 1735f-7a (2000) (preempting all state usury laws imposed on loans secured by first mortgages on residential real estate). Thanks to the exportation doctrine, however, the DIDMCA's preemptive effect shields federally insured credit card lenders. *See* THE COST OF CREDIT, *supra* note 61, at 70-71.

99. 12 U.S.C. §§ 3801, 3803 (2000).

100. Marquette Nat'l Bank of Minneapolis v. First of Omaha Serv. Corp., 439 U.S. 299 (1978).

101. *Id.* at 302 n.4.

102. One commentator writes:

> [T]he most distinguishing feature of the credit card industry is its recent concentration.... Today, the top ten card issuers control nearly 77 percent of the card market and 69 percent of the $1.2 trillion in 1999 charge volume.... Indeed, industry mergers and acquisitions have reduced the number of major corporate players and thus contributed to the costly increase in consumer banking services in general . . . and credit card finance charges and fees in particular.

MANNING, *supra* note 4, at 95.

103. *See supra* notes 58, 60. Michael Barr attributes the decreased availability of small, unsecured loans from banks to this deregulatory trend. Michael S. Barr, *Banking the Poor,* 21 YALE J. ON REG. 121, 152 (2004). For perspectives on deregulation of the banking industry, see, e.g., KERRY COOPER & DONALD R. FRASER, BANKING DEREGULATION AND THE NEW COMPETITION IN FINANCIAL SERVICES (Student ed. 1986); THE DEREGULATION OF THE BANKING AND SECURITIES INDUSTRIES (Lawrence G. Goldberg & Lawrence J. White, eds., Lexington Books 1979); ALAN GART, REGULATION, DEREGULATION, REREGULATION: THE FUTURE OF THE BANKING, INSURANCE, AND SECURITIES INDUSTRIES (John Wiley & Sons, Inc. 1994); Helen A. Garten, *Regulatory Growing Pains: A Perspective on Bank Regulation in a Deregulatory Age,* 57 FORDHAM L. REV. 501 (1989); Jith Jayaratne & Philip E. Strahan, *Entry Restrictions, Industry Evolution, and Dynamic Efficiency: Evidence from Commercial Banking,* 41 J.L. & ECON. 239 (1998); Arthur E. Wilmarth, Jr., *The Transformation of the U.S. Financial Services Industry, 1975-2000: Competition, Consolidation, and Increased Risks,* 2002 U. ILL. L. REV. 215 (2002).

104. For a detailed discussion of the regulatory efforts to equalize competition between state and federally chartered banks, see Schiltz, *supra* note 7, at 540-56.

The "exportation doctrine" developed in *Marquette* is a primary example.[105] It has been continually expanded by federal courts, the Office of the Comptroller of the Currency (OCC), and the Office of Thrift Supervision (OTS) to include a broad range of lenders[106] and to limit the effect of state consumer protection laws regulating penalties,[107] terms of credit,[108] and disclosure.[109]

Because usury caps were among the few state laws imposing substantive limitations on credit cards, their preemption allowed much greater exploitation by credit card issuers. The result, according to one empirical study, was the rapid expansion of the credit card industry, eighty percent of whose profits derive from interest payments.[110] By eliminating the major substantive limitations on credit card lenders, the law certainly increased availability of credit to a broader range of borrowers.[111] As credit card borrowers incurred vast amounts of debt, comparable only to mortgage debt, the law also left them with very few substantive rights other than the right to information.

As Part II.A. describes, courts have been extraordinarily reluctant to void or reinterpret credit card contracts in favor of borrowers. In marked contrast to the mortgage lending context, courts have upheld arbitration clauses,[112] penalty

---

105. *Id.* at 546-47.

106. *Id.* at 535, 561-65, 602; *see supra* note 60.

107. *See, e.g.*, Smiley v. Citibank, 517 U.S. 735, 739 (1996) (preempting state regulation of late payment fees); Greenwood Trust Co. v. Massachusetts, 971 F.2d 818, 830 (1st Cir. 1992) (preempting state regulation of late fees); Watson v. First Union Nat'l Bank, 837 F. Supp. 146, 150 (D.S.C. 1993) (preempting state regulation of credit card overlimit fees); Hill v. Chem. Bank, 799 F. Supp. 948, 953 (D. Minn. 1992) (citing precedent to preempt state regulation of overlimit fees); Tikkanen v. Citibank, 801 F. Supp. 270, 278-79 (D. Minn. 1992) (preempting state regulation of both overlimit and late payment fees); Stoorman v. Greenwood Trust Co., 908 P.2d 133, 135 (Colo. 1995) (preempting state regulation of late payment fees); Copeland v. MBNA Am. Bank, 907 P.2d 87, 94 (Colo. 1995) (preempting state regulation of late payment fees).

108. *See* Schiltz, *supra* note 7, at 615-16.

109. Am. Bankers Ass'n v. Lockyer, 239 F. Supp. 2d 1000, 1013-14 (E.D. Cal. 2002) (accepting the argument made by plaintiffs and OCC as amicus curiae that § 85 preempted a California statute imposing disclosure requirements on credit card lenders that were more stringent than TILA).

110. SULLIVAN ET AL., *supra* note 5, at 135 (also noting that the remaining 20% of profits come from "annual fees, late fees, over-limit fees, and merchant fees"). *See also* Lawrence M. Ausubel, *Credit Card Defaults, Credit Car Profits, and Bankruptcy*, 71 AM. BANKR. L.J. 249 (1997).

111. Professors Sullivan, Warren and Westbrook state: "The enormous profits available from people who charge up to the limits and pay only the minimum each month have made delinquent card holders who pay high interest the most valued customers in the business. . . . The credit card companies hope to expand credit card debt even further. . . . Their most aggressive targeting is aimed at the only group left that is not already inundated with plastic: the poor." *Id.* at 135-36.

112. *See, e.g.*, Taylor v. First N. Am. Nat'l Bank, 325 F. Supp. 2d 1304 (M.D. Ala. 2004); Arriaga v. Cross Country Bank, 163 F. Supp. 2d 1189 (S.D. Cal. 2001); Bank One, N.A. v. Coates, 125 F. Supp. 2d 819 (S.D. Miss. 2001); Edelist v. MBNA Am. Bank, 790 A.2d 1249

clauses, and onerous pricing terms,[113] often on grounds of preemption and in the face of compelling parol evidence that such terms were inappropriately imposed. In so doing, courts have established that the only meaningful backstops to exploitative lender behavior, other than the disclosure regime of the CCPA, are traditional contract defenses, such as unconscionability, fraud, and bad faith.[114]

While mortgage lenders today are subject to the equity of redemption, limitations on foreclosure and other penalties, judicial scrutiny of contract formation and interpretation, and, depending on the type of lending institution, some usury caps,[115] credit card lenders are subject to very few consumer protection regulations other than the CCPA. While the disclosure regime of the CCPA supplements the more substantive property protections provided by traditional mortgage law to residential mortgage borrowers, credit card borrowers must rely extensively on disclosure for protection.

C. Homestead and Other Exemptions and Bankruptcy

There are two important exceptions to the disclosure-only regime in credit card borrowing. While powerful, however, these protections do not ultimately neutralize the important differences between the two legal regimes for consumer borrowers.

The first group of protections are state homestead and personal property exemptions, which exempt certain items from seizure by judgment creditors.[116] In some states, these laws generously define the necessities of life, leaving a relatively broad range of individual debtors judgment-proof.[117] If a credit card or other unsecured lender wins a judgment against a borrower, that lender may nonetheless be unable to collect.[118] A mortgagee, on the other hand, is a secured creditor who can collect the debt owed by foreclosing on the collateral. Such a remedy does not exempt the borrower's house when it is used as

---

(Del. Super. Ct. 2001). The vast majority of reported cases in which consumers have challenged specific clauses of credit card contracts have avoided consideration of such claims on the grounds that a valid arbitration clause precluded such consideration.

113. *See, e.g.*, Morris v. Redwood Empire Bancorp, 27 Cal. Rptr. 3d 797, 809 (Cal. Ct. App. 2005) (holding that $150 fee charged by credit card company to terminate a "merchant account" that had been opened by an elderly disabled man with a home-based small business was not unconscionable). This appears to be one of the few cases not dismissed on grounds of preemption or an applicable arbitration clause.

114. *But see infra* Part I.C.

115. *But see supra* note 61 regarding the expansion of the exportation doctrine to cover a broader range of lenders.

116. For an overview of exemption law, see CONSUMER CREDIT LAW MANUAL, *supra* note 8, § 9; NATIONAL CONSUMER LAW CENTER, FAIR DEBT COLLECTION § 12 (5th ed. 2004) [hereinafter FAIR DEBT COLLECTION].

117. *See* FAIR DEBT COLLECTION, *supra* note 116, Appendix F, at 751-78.

118. Limitations on attachment of wages and bank accounts also protect borrowers. *See* CONSUMER CREDIT LAW MANUAL, *supra* note 8, § 9.

collateral, no matter how much of a necessity that house is.[119] If none of the legal protections described above prevent foreclosure by the mortgagee, a mortgagor will ultimately lose her home. In other words, while the credit card borrower may have fewer protections on the front end (before she is adjudicated to be in default), she does have a significant protection at the point of loss of personal assets.

The second important consumer protection is bankruptcy. Bankruptcy law provides special protections to residential mortgagees. Specifically, debtors filing for bankruptcy under Chapter 13 are limited in their ability to modify debt secured by the "debtor's principal residence," thereby advantaging residential mortgagees over other secured creditors.[120] This legal benefit for mortgage lenders has, until recently, counseled in favor of incurring credit card debt, which could be fully discharged,[121] over residential mortgage debt, which could not even be substantially modified, where bankruptcy is a serious possibility. The new Bankruptcy Abuse Prevention and Consumer Protection Act of 2005[122] has changed this picture. This statute makes it more difficult for consumers to discharge credit card debt,[123] thereby putting credit card lenders on more equal footing with mortgage lenders. Under the new bankruptcy law, credit card borrowers may no longer benefit from as much protection as they once had.

While important, these two major protections occur late in the process of a typical credit dispute. It may be true that a credit card borrower who has lost her day in court will be less likely to lose her home than a mortgagor who has lost her day in court. But it also appears true that a credit card borrower is much more likely to lose her day in court than is a mortgagor.[124] Moreover, these two protections are only available at the point of catastrophe, and homestead and other exemptions are most effective for the poorest borrowers. While progressive in important respects, these protections are far from satisfactory for the average middle-class consumer. Certainly, they do not replace the diffuse borrower sensitivity existing in mortgage law even today.

---

119. Mortgagees are typically excluded from homestead exemptions. *See* FAIR DEBT COLLECTION, *supra* note 116, at 508. In Florida, which exempts a homestead of up to 160 acres, an individual would be well-advised to borrow on something other than her house. By choosing credit card over home equity debt, for example, she would be able to exempt her home from attachment by credit card lenders, while not having to worry about foreclosure by a mortgagee.

120. 11 U.S.C § 1322(b)(2) (2000).

121. *See* 11 U.S.C. §§ 727, 1328 (2001).

122. Bankruptcy Abuse Prevention and Consumer Protection Act of 2005, Pub. L. No. 109-8, 119 Stat. 23 (primarily codified in scattered sections of 11 U.S.C.).

123. *See* Joel Marker, *Get Ready for the Bankruptcy Amendments of 2005*, 18 UTAH B. J. 12, 13 (2005) (describing ways in which the new statute limits discharge of credit card debt).

124. Here, I am referring to the mortgagor's day in court prior to a foreclosure proceeding.

D. Two Paradigms, Side by Side

By their very nature, the status-based substantive protections developed during the regulatory phase of mortgage law had the purpose of protecting the borrower's right to his collateral. It is not surprising, then, that these protections are quintessential property rules. They operate much like a specific performance remedy by keeping the collateral in the borrower's hands rather than providing a monetary substitute.[125] They form a superstructure over the private bargaining of the lender and borrower and are inherently local in nature because their focus is property in a particular geographic location. Moreover, a notable number of rules in the mortgage context are typically non-waivable.[126]

The absence of land from the credit card transaction might have made the borrowers' protections developed by mortgage law a counter-intuitive model for credit card law. After all, what use would an equity of redemption be to a credit card borrower? Added to the absence of collateral was the nationalization of the banking industry, a phenomenon which deregulation facilitated, at the very moment of the credit card's ascendance. Whether for these reasons or others, the model of regulation that was chosen was contract-based, where parties are free to bargain for rights, remedies are awarded in the form of damages, and the backstops to exploitation are the limited defenses of unconscionability and fraud.

In this respect, however, legal and consumer perspectives diverge. For the borrower, the choice between borrowing regimes is not so dramatic. Credit limits on credit card accounts can be as high as home equity loans.[127] Teaser interest rates on credit cards can be as low as or lower than interest rates on

---

125. This is the classic definition of a property rule, as discussed by Calabresi and Melamed in the context of nuisance law. In their words, "No one can take the entitlement to private property from the holder unless the holder sells it willingly and at the price at which he subjectively values the property." Calabresi & Melamed, *supra* note 53, at 1105. Their definition of a liability rule is one where "an external, objective standard of value is used to facilitate the transfer of the entitlement from the holder . . . ." *Id.* at 1106.

126. The inalienable nature of some rules puts them in a third category described by Calabresi and Melamed, that of inalienability rules: "[T]he law not only decides who is to own something and what price is to be paid for it if it is taken or destroyed, but also regulates its sale—by, for example, prescribing preconditions for a valid sale or forbidding a sale altogether." Calabresi & Melamed, *supra* note 53, at 1111. For purposes of my comparison of property and contract rules, the non-waivability of these rules makes them more propertarian in nature.

127. Consider these advertisements from major credit card companies: MBNA, http://www.mbna.com/loans/term_line.html (last visited Sept. 12, 2006) (advertising a card with a credit line up to $25,000 "to use for virtually any large purchase - build a pool, buy a computer, upgrade your home appliances, and more," for which no collateral is required); Discover Credit Cards, http://www.creditcards.com/Discover.php (last visited Nov. 14, 2005) (advertising a platinum card with up to a $50,000 credit limit); Advanta Cards, http://www.credit.com/products/credit_cards/advanta.jsp (last visited Sept. 12, 2006) (advertising a platinum rewards card with a "large credit limit").

home equity loans.[128] Either form of loan can be closed-end or open-end.[129] Middle class and lower income individuals typically have easy access to both types of loans[130] and regularly use both forms of borrowing to make ends meet.[131]

Despite the different protections available to secured and unsecured creditors, the losses to consumers can be equally devastating in either regime.[132] The credit card creditor must obtain a judgment in order to acquire a lien. For the borrower, however, the consequences, namely loss of physical assets, could be quite similar.[133]

Moreover, the substantive protections developed during the regulatory phase of mortgage law do not exclusively protect the borrower's right to his collateral. As this section and the next show, these protections have produced a more equal loan transaction. For example, the equity of redemption overrides strict time limits on payment. Foreclosure typically provides meaningful notice to the borrower of loss of rights, and courts regularly void onerous loan terms as antithetical to the borrower's intentions. Thus, the average middle-class mortgage borrower today typically has better loan terms than the average middle-class credit card borrower.[134]

II. THE CONSEQUENCES

Part I suggests that the creation of a lending structure ungrounded in real assets has far-reaching consequences for the consumer borrower. The remainder of this Article explores these consequences.

---

128. For example, see Wells Fargo, http://www.wellsfargo.com/credit_cards/select_card/ (last visited Sept. 12, 2006) (offering cards with introductory rates as low as 0%); Low Interest Cards, http://www.creditcards.com/low-interest.php (last visited Nov. 11, 2005) (offering cards with introductory rates of 0% from various credit card companies); CitiBank Cards, http://www.credit.com/products/credit_cards/citi.jsp (last visited Sept. 12, 2006) (offering a card for college students with an introductory rate of 0%); Chase Cards, http://www.credit.com/products/credit_cards/chase.jsp (last visited Nov. 11, 2005) (offering a card for students with an introductory rate of 0%); Discover Credit Cards, http://www.credit.com/products/credit_cards/discover.jsp (last visited Sept. 12, 2006) (offering cards with introductory rates as low as 0%, including for college students).

129. ROHNER & MILLER, *supra* note 4, at 245-48; THE COST OF CREDIT, *supra* note 61, at 366-69.

130. *See supra* note 14 and accompanying text.

131. It is unquestionably true that federal tax policy, namely the deductability of mortgage interest, and other factors make residential mortgage lending arguably more attractive. *See generally* Forrester, *supra* note 13. These factors are often counterbalanced by the ease and seductiveness of credit card borrowing. *See generally* Bar-Gill, *supra* note 5.

132. *See supra* Part I.C.

133. *But see supra* Part I.C.

134. I am not referring here to predatory mortgage lending, where loan terms are much more onerous. See *infra* Part II.B. for a discussion of predatory lending.

A. In Court

This section closely examines the significantly different judicial treatment of the two types of borrowers, detailing the ways in which courts have shaped the two paradigms and demonstrating more fully that a borrower's choice of borrowing regime profoundly impacts her legal rights and protections. As the cases in this section demonstrate, a borrower's day in court is a cumulative result of the prototype by which she is defined. These cases[135] represent the norm rather than the exception.[136]

1. Mortgages

In *McGill v. Biggs*, Plaintiff John McGill wanted to provide a proper funeral for his mother.[137] After unsuccessfully trying to obtain a $1,500 bank loan, presumably unsecured, he turned to his half-uncle, Defendant Johnny Biggs, for help.[138] According to Plaintiff, Biggs agreed to give him the funds in return for "collateral" in the form of a quit-claim deed in a house which Plaintiff had jointly owned with his mother.[139] Plaintiff made one payment to Biggs but then made no other payments for more than two years, at which point he offered to repay Biggs in full, an offer which Biggs refused.[140] Six months later, Biggs recorded the deed and tried to list it with a broker as a rental property.[141] Plaintiff sued, claiming that he was the owner of the house, and that the deed merely evidenced a mortgage agreement.[142] Despite the recorded deed in favor of Defendant and the apparent absence of any other writing evidencing a mortgage agreement, both the trial court and appellate court ruled in Plaintiff's favor, effectively rejecting the proposition that a contract for conveyance was ever formed.[143] They did so in the absence of a finding of

---

135. The mortgage case is cited in a well-known treatise, as well as in the Restatement. *See* REAL ESTATE FINANCE LAW, *supra* note 19, at 57-58; RESTATEMENT (THIRD) OF PROPERTY (MORTGAGES) §§ 3.2 (1997). The credit card case is among the cases listed by one author as characteristic of courts' treatment of cases with similar facts. Bar-Gill, *supra* note 5, at 1424 n.213.

136. The mortgage case is not representative of what happens in a foreclosure proceeding. Once a borrower is that far along in the post-default process, courts appear less inclined to reinterpret the mortgage agreement in favor of the borrower and less inclined to overturn the foreclosure sale in favor of protecting the borrower. *See supra* note 39.

137. McGill v. Biggs, 434 N.E.2d 772, 773 (Ill. App. Ct. 1982).

138. *Id.*

139. *Id.* Plaintiff was the surviving joint tenant in the house. *Id.* at 772-73. Because many mortgages today have language providing fee title in the collateral to the lender, McGill's assumption that he was only providing collateral for a loan was not completely unfounded.

140. *Id.* at 773.

141. *Id.*

142. *Id.*

143. *Id.* at 775. Many courts have held that a deed, though signed by only the grantor, has the effect of a signed contract, thereby typically deserving application of the parol evidence rule

either fraud or unconscionability and apparently without heed to the Statute of Frauds or the Parol Evidence Rule.

What could have justified such a surprising result? An Illinois statute provided that a "deed conveying real estate, which shall appear to have been intended only as a security . . . shall be considered as a mortgage."[144] According to Illinois case law, such a determination can be made on the basis of "almost every conceivable fact that could legitimately aid" it.[145] In this case, the court considered the facts that McGill owed Biggs money and that both parties believed that the debt remained unpaid, that he had been unable to obtain funds from other sources, that the parties were in a close relationship, that McGill was not represented by an attorney in the transaction, that he continued to treat the house as his own, that he did not intend to sell the house to Biggs, and that there was a large disparity between the loan amount and the value of the house.[146] Both the trial and appellate courts considered the context of the transaction.

Although the appellate opinion did not describe it as such, the *McGill* holding effectively restored the equity of redemption to Plaintiff. While cases interpreting conveyances as mortgages are the most dramatic, such contextual treatment by courts is the rule in mortgage transactions.[147] Courts have repeatedly voided remedies, such as strict foreclosure, options to purchase, and deeds in lieu of foreclosure, in order to protect the equity of redemption.[148]

---

and other contract doctrines. *See, e.g.*, *In re* Weitzel's Estate, 280 N.W. 270, 272 (Neb. 1938) ("In so far as a deed is deemed to be contractual, it is subject to the parol evidence rule the same as any other contract." (quoting 4 NEB. L. BULL. § 11, 134 (1926))); Rael v. Cisneros, 487 P.2d 133, 136 (N.M. 1971) (a deed is "a specialized form of contract"); Radspinner v. Charlesworth, 369 N.W.2d 109, 112 (N.D. 1985) ("A deed is a written contract, and is subject to the [parol evidence] rule."); John Deere Indust. Equip. Co. v. Gentile, 459 N.E.2d 611, 614 (Ohio Ct. App. 1983) ("The parol evidence rule excludes evidence of prior or contemporaneous statements contradicting or varying the terms of a written deed or contract, which is complete and unambiguous on its face."). The recording of a deed raises its utility as evidence of ownership to an even higher level. *See generally* Corwin W. Johnson, *Purpose and Scope of Recording Statutes*, 47 IOWA L. REV. 231 (1962).

144. *McGill*, 434 N.E.2d at 773 (quoting ILL. REV. STAT. 1977, ch. 95, par. 55). Such statutes regularly supplement case law to the same effect. *See* REAL ESTATE FINANCE LAW, *supra* note 19, at 50-51.

145. *McGill*, 434 N.E.2d at 773. Illinois courts have developed a list of circumstances extrinsic to the deed that ought to be considered, many of which the court did consider in this case. *Id.* at 774.

146. *Id.* at 773-75.

147. The Restatement of Mortgages has institutionalized this doctrine. RESTATEMENT (THIRD) OF PROP.: MORTGAGES §§ 3.2, 3.3 (1997).

148. First Illinois Nat'l Bank v. Hans, 493 N.E.2d 1171, 1175 (Ill. Ct. App. 1986) (voiding a deed-in-lieu of foreclosure); Humble Oil & Ref. Co. v. Doerr, 303 A.2d 898, 916 (N.J. Super. Ct. Ch. Div. 1973) (finding the option to purchase "unenforceable"); Basile v. Erhal Holding Corp., 538 N.Y.S.2d 831, 833 (App. Div. 1989) (voiding the deed-in-lieu of foreclosure); Dawson v. Perry, 30 Va. Cir. 372, 376 (1993) (finding the deed-in-lieu of foreclosure invalid);

State statutes regularly supplement case law by, among other things, limiting late fees that can be charged[149] to protect the debtor's ownership rights. In this case, McGill's testimony that he did not understand what a "quit-claim" deed was, that he thought he was signing something he was not, and that he did not have an attorney were all facts supporting the voiding of the contract.[150] In this respect, the court's language and imagery are quite telling. The court variously described the borrower as "vague as to the nature of the transaction,"[151] "upset and distressed when the transaction took place,"[152] and "falling on hard times by drinking heavily."[153] It noted that the "disparity of the situations of the parties" was one circumstance to be considered in determining whether the transaction was in fact a mortgage.[154]

2. Credit Cards

In *Arriaga v. Cross Country Bank*, Plaintiff Andrea Arriaga[155] responded to a promotional mailing from Defendant Cross Country Bank (CCB)[156] by submitting an application for a CCB credit card.[157] In her class action lawsuit against CCB in federal district court, Arriaga claimed that before she received her credit card, she was charged "$150 in up-front fees," which CCB had not disclosed to her.[158] Shortly after she received her credit card, CCB charged her for an "Applied Advantage" membership, which she had neither known about nor requested.[159] Plaintiff's theory was that these charges were part of a fraudulent scheme by CCB to force customers over their credit limits, thereby

---

Barr v. Granahan, 38 N.W.2d 705, 708 (Wis. 1949) (determining "that the purpose of the option agreement was to secure the mortgage indebtedness"); *see also* John C. Murray, *Clogging Revisited*, 33 REAL PROP. PROB. & TR. J. 279 (1998) (discussing deeds-in-lieu of foreclosure as clogs on the equity of redemption).

149. REAL ESTATE FINANCE LAW, *supra* note 19, at 449-50. Note, however, that some of these statutes are preempted by the National Bank Act and OTS regulations. *Id.* at 451.

150. *McGill*, 434 N.E.2d at 773.

151. *Id.*

152. *Id.*

153. *Id.*

154. *Id.* at 774.

155. Plaintiff sued on behalf of herself and a similarly situated class. Arriaga v. Cross Country Bank, 163 F. Supp. 2d 1189, 1190 (S.D. Cal. 2001). For simplicity's sake, I will refer to all plaintiffs as the plaintiff.

156. Plaintiff also named Applied Card Systems, Inc., which promoted the Applied Advantage program on behalf of CCB. Id. For simplicity's sake, I will refer to both as the defendant.

157. *Id.*

158. *Id.*

159. *Id.* at 1190-91. She also claimed that CCB "purposefully failed to send monthly billing statements, forcing her and other customers to request one at a charge of $3 to their credit accounts." *Id.* at 1191.

causing them to incur "over limit" fees.[160] She claimed violations of TILA, California deceptive trade practices and unfair competition statutes, fraud, and breach of contract.[161]

On an abstract level, the facts of the Arriaga and McGill cases are comparable. Both involve consumer debt, with a lender more sophisticated than the borrower. Both have suggestions of fraudulent behavior, in which the plaintiff believed he or she was entering into a transaction that the lender defined differently. Moreover, the facts of Arriaga are arguably more egregious than McGill. In McGill, for example, Plaintiff had failed for more than two years to make any payments to Defendant for the alleged loan, while in *Arriaga*, there was no evidence of late payment or any other default by Plaintiff.[162] A key difference in the cases' facts is that McGill's loan was secured by a house.[163]

But both the results and the analyses of the two cases were so different that it seems almost illogical to compare them. While the *McGill* court considered the broad context surrounding the transaction in order to grant Plaintiff relief,[164] the *Arriaga* court kept strictly to the terms of the credit card contract.[165] In doing so, it rejected Arriaga's arguments for voiding the arbitration clause in Arriaga's credit card contract[166] and instead granted Defendant's motion to compel arbitration.[167] Arriaga's argument that the clause had been obtained by fraudulent inducement and was, therefore, unconscionable was among the most persuasive.[168] Arriaga contended that she was "forced to agree" with the arbitration clause and with the credit card agreement as a whole because CCB had already charged her account by the time it sent her the credit card agreement.[169] Thus, Arriaga would have risked incurring a negative credit report by contesting the terms of the agreement.[170] The court, however, refused to consider these facts because Supreme Court precedent[171] limited it only to deciding whether the arbitration clause, independent of the remainder of the

160. *Id.* at 1191.

161. *Id.* at 1190.

162. *Id.*; *McGill*, 434 NE.2d at 772-73.

163. *McGill*, 434 N.E.2d at 772-73.

164. *Id.* at 774-75.

165. *Arriaga*, 163 F. Supp. 2d at 1191.

166. In some of her arguments, plaintiff asked the court not to apply the arbitration clause to some of her claims, rather than void it altogether. *Id.*

167. The court granted this motion and stayed the court proceeding pending arbitration. *Id.* at 1202.

168. *Id.* at 1193.

169. *Id.* The court also rejected her argument that some of her claims arose before she had entered into the credit card agreement with the arbitration clause on the grounds that the language of the arbitration clause covered all claims that "'relate to or arise out of'" the credit card agreement. *Id.* at 1192 (quoting Arriaga's credit card agreement) (emphasis omitted).

170. *Id.* at 1193.

171. Prima Paint Corp. v. Flood & Conklin Mfg. Co., 388 U.S. 395, 399-400 (1967).

contract, was fraudulently obtained or unconscionable.[172] The court did this despite the Federal Arbitration Act's broad statement that an arbitration clause shall be valid "save upon such grounds as exist in law or in equity for the revocation of any contract."[173] Regarding the facts related to the arbitration clause, the court held that none of these were so egregious as to be fraudulent or unconscionable.[174] The court found that Arriaga was enlightened enough to understand the terms that she was accepting in the agreement and to understand that she could opt out of them.[175] Here, too, the court's language and imagery are telling. In denying Arriaga's challenge to the arbitration clause as unconscionable, the court stated that the terms were not "beyond a reasonable persons [sic] expectations,"[176] in part because the contract "clearly states the [National Arbitration Forum's] toll-free number and world wide web address from which Arriaga could have obtained the full terms and rules in question."[177]

Could this case have looked more like *McGill* if the court had not interpreted precedent to require such a narrow review of only that clause rather than the whole contract or if there had been no arbitration clause at all? One way to answer this question is to consider cases in which courts did void such clauses. The few credit card cases that have done so have looked at a broader range of circumstances to void the arbitration clause, including the circumstances surrounding the formation of the contract.[178] In doing so,

---

172. *Arriaga*, 163 F. Supp. 2d at 1193.

173. 9 U.S.C. § 2 (2005).

174. *Arriaga*, 163 F. Supp. 2d at 1193-95.

175. *Id.* at 1194.

176. *Id.*

177. *Id.* This type of reasoning is ubiquitous in credit card cases interpreting arbitration clauses. *See supra* note 112.

178. An example of a case considering the circumstances surrounding contract formation that has subsequently been positively cited in both federal and state courts said:

> The [lower] court's focus on the [alternative dispute resolution (ADR)] clause, standing alone, was misplaced: it is the Bank's exercise of its discretionary right to change the agreement, not the ADR clause in and of itself, which must first be analyzed in terms of the implied covenant. If the Bank's performance under the change of terms provision was not consonant with the duty of good faith and fair dealing, then whether the ADR clause, considered in isolation, satisfies the implied covenant makes no difference.

Badie v. Bank of Am., 79 Cal. Rptr. 2d 273, 284 (Cal. Ct. App. 1998).

In 2004, a federal district court followed *Badie*'s precedent stating:

> The Court agrees with the *Badie* court that the terms discussed in the change-in-terms clause must supply the universe of terms which could be altered or affected pursuant to the clause. To hold otherwise would permit the Bank to add terms to the Customer Agreement without limitation as to the substance or nature of such new terms. . . . Thus, the arbitration clause does not bind plaintiff.

Stone v. Golden Wexler & Sarnese, P.C., 341 F. Supp. 2d 189, 198 (E.D.N.Y. 2004); Discover Bank v. Shea, 827 A.2d 358, 361 (N.J. Super. Ct. Law Div. 2001) (also citing *Badie* for the same proposition).

however, these courts were still largely confined to common law defenses such as lack of good faith[179] and unconscionability.[180] The courts did not consider the breadth of context that is typical in cases such as *McGill*. These claims are also very difficult to win because they require a showing of egregious behavior by defendants.[181] Moreover, the crux of these analyses was whether the lenders provided sufficient disclosure to accomplish contract formation. Courts examined whether lenders had a unilateral right to change the credit card agreement,[182] whether the language was ambiguous,[183] and whether the cardholders had opportunities to request more information.[184]

Perhaps most importantly, even where an arbitration clause does not deprive a cardholder of her day in court, the cardholder's claims are far less powerful than those available to a mortgagor. Arriaga's claims covered the entire field available to her, yet her bases for relief were limited to claims of inadequate disclosure, pursuant to TILA and other sections of the CCPA, and state law claims for unfair competition, deceptive practices, fraud, and unconscionability.[185] The federal claims focus on the question of whether the defendant properly disclosed its onerous loan terms, without examining the onerous nature of the terms themselves. Meanwhile, state law claims, such as these, typically require very high thresholds of improper behavior to warrant relief.[186] When remedies are granted, litigation is costly and uncertain due to

---

179. *Stone*, 341 F. Supp. 2d at 197-98; *Badie*, 79 Cal. Rptr. 2d at 284.

180. Nefores v. Branddirect Mktg., Inc., No. 03-CA-104, 2004 WL 2260703, at *5 (Ohio Ct. App. Sept. 17, 2004); *Shea*, 827 A.2d at 365-68.

181. *See, e.g.*, *Arriaga*, 163 F. Supp. 2d at 1195 ("[C]ourts have held that in order for a contract term to be substantively unconsionable, it must be so one-sided as to 'shock the conscience.'" (citation omitted)).

182. *Shea*, 827 A.2d at 365; *Badie*, 79 Cal. Rptr. 2d at 281-84.

183. *Badie*, 79 Cal. Rptr. 2d at 286-89.

184. *Arriaga*, 163 F. Supp. 2d at 1194-95.

185. *Id.* at 1190.

186. The discussions of unconscionability in the credit card cases upholding arbitration clauses certainly evidence a difficult threshold. *See* Siemens Credit Corp. v. Newlands, 905 F. Supp. 757, 765 (N.D. Cal. 1994) ("A (substantively) unconscionable bargain has been defined as one that 'no man in his senses and not under delusion would make on the one hand, and no honest and fair man would accept on the other.'") (quoting Hume v. United States, 132 U.S. 406 (1889)). This standard has not changed in the intervening 100 years. *See, e.g.*, Garrett v. Janiewski, 480 So. 2d 1324, 1326 (Fla. Dist. Ct. App. 1985) (stating that the word unconscionable is synonymous with "shocking to the conscience" and "monstrously harsh"). The court noted that proof of unconscionability typically requires both a showing of substantive and procedural unconscionability. *Id.* Fraud has always required a high threshold of proof. *See, e.g.*, Wärtsilä NSD N. Am., Inc. v. Hill Int'l, Inc., 342 F. Supp. 2d 267, 274 (D. N.J. 2004) ("It is in the nature of fraud cases that the scienter element will ordinarily be difficult to establish."). For an analysis of the challenges involved in litigating unfair and deceptive practices claims on behalf of consumers, see Anthony Paul Dunbar, Comment, *Consumer Protection: The Practical Effectiveness of State Deceptive Trade Practices Legislation*, 59 TUL. L. REV. 427 (1984).

the highly fact-specific nature of these defenses.[187] Moreover, even the broadest state unfair competition laws require an unfair act by the defendant.[188] These claims, some of which are regularly preempted by broad interpretations of the exportation clause,[189] provide significantly less protection than property rules, such as the equity of redemption, which requires no bad act by the lender and applies even in the face of defaulting behavior by the borrower.[190]

Alternatively, the question whether *Arriaga* could have looked more like *McGill* could be moot because many credit card contracts have arbitration clauses. Given the rule of deference to arbitration,[191] it is reasonable to expect that in the vast majority of credit card cases, arbitrators will decide the merits of cardholders' claims of substantive unfairness. Statistics concerning outcomes do not favor cardholders.[192] Meanwhile, in the mortgage context, statutes govern the important creditors' remedy of foreclosure, not arbitration clauses.[193]

---

187. *See* Shanker, *supra* note 19, at 90-93 (discussing the limitations of the unconscionability doctrine in protecting mortgagors); Jean R. Sternlight & Elizabeth J. Jensen, *Using Arbitration to Eliminate Consumer Class Actions: Efficient Business Practice or Unconscionable Abuse?*, 67 LAW & CONTEMP. PROBS. 75 (2004) (discussing the limitations of unconscionability in preventing businesses from eliminating class action rights of consumers); Russell Korobkin, *Bounded Rationality, Standard Form Contracts, and Unconscionability*, 70 U. CHI. L. REV. 1203 (2003) (proposing modifications to the unconscionability doctrine to better protect buyers).

188. *See, e.g.*, ALA. CODE §§ 8-19-1 to -15 (Supp. 1983); ALASKA STAT. §§ 45.50.471 to .561 (1983); COLO. REV. STAT. §§ 6-1-101 to -114 (1973).

189. *See supra* notes 107-109.

190. An example of an important exception is *Discover Bank v. Owens* where the court denied a bank's collection action because:

> Even if plaintiff was technically within its rights in its handling of defendant's account, it was unreasonable and unjust for it to allow defendant's debt to continue to accumulate well after it had become clear that defendant would be unable to pay it. . . . This court has broad legal and equitable powers, and now brings them to bear for the debtor in this case.

Discover Bank v. Owens, 822 N.E.2d 869, 873-75 (Ohio Misc. 2d 2004). While breathtaking in the credit card context, this type of language is typical in the mortgage context.

191. This is the rule where credit card agreements incorporate the provisions of the Federal Arbitration Act, 9 U.S.C. §§ 1-14 (2005).

192. *See* Jean R. Sternlight, *Creeping Mandatory Arbitration: Is it Just?*, 57 STAN. L. REV. 1631, 1658 (2005) ("[O]ne credit card company's statistics showed that in the two years after it imposed mandatory arbitration, only four customers filed claims against the company. In contrast, during that same period the company brought 51,622 arbitration claims against consumers."); Caroline E. Mayer, *Win Some, Lose Rarely? Arbitration Forum's Rulings Called One-Sided*, WASH. POST, Mar. 1, 2000, at E1 (reporting that First USA won 99.6% of its cases before a National Arbitration Forum arbitrator between January 1998 and November 1999); *see also* Public Citizen, *Mandatory Arbitration Clauses: Undermining the Rights of Consumers, Employees, and Small Business*, http://www.citizen.org/congress/civjus/arbitration/articles.cfm?ID'7332 (critiquing mandatory arbitration clauses); Remar Sutton, *The Big Problem*, http://www.givemebackmyrights.com/bma-problem.htm (discussing how mandatory arbitration clauses hurt consumers). *But see* ADR News, *Credit Card Holders Not Disadvantaged in*

While the *McGill* court responded to the borrower's ignorance about his contractual rights by filling in the picture outside the contract,[194] the *Arriaga* court found a knowledgeable waiver of rights by the borrower in the absence of fraud or unconscionability.[195] While the *McGill* court concluded that the writing evidencing the parties' transaction did not represent the actual agreement of the parties,[196] the *Arriaga* court assumed that it did.[197] And while the *McGill* court restored a right to the borrower for which he did not explicitly negotiate,[198] the *Arriaga* court limited the borrower's rights to those evidenced in a writing, which had undisputedly not been negotiated.[199] The result of cases such as these is that ignorant mortgagors have more opportunity to have their fully contextualized day in court, while enlightened credit card borrowers are expected to understand the consequences of agreeing to arbitration and other onerous terms.[200]

B. In the Marketplace: Predatory Credit Card Lending

While it may be unexpected, the stark incongruity in legal treatment of the two types of borrowers described in Part II.A. is well documented. The incongruity in the two credit markets, which this section explores, is less well documented. The phenomenon of predatory lending is integral to this exploration.

Although predatory lending is an established concept in the mortgage market,[201] it has barely been identified in the credit card market.[202] In the

---

Arbitration, Study Says, 60 DISP. RESOL. J. 5 (2005) (finding that "consumers in NAF arbitrations prevailed more often than businesses in cases that went to an arbitration hearing (55% of the cases were resolved in the consumers' favor)").

193. In non-judicial foreclosure states there is much less process for borrowers, thus, making those foreclosure cases look more like credit card arbitrations.

194. McGill v. Biggs, 434 N.E.2d 772, 774-75 (Ill. App. Ct. 1982).

195. The court only applied the doctrines of fraud and unconscionability to the question of whether the arbitration clause was appropriately negotiated. Arriaga v. Cross Country Bank, 163 F. Supp. 2d 1189, 1193-95 (S.D. Cal. 2001).

196. *McGill*, 434 N.E.2d at 775.

197. *Arriaga*, 163 F. Supp. 2d at 1192-93.

198. *McGill*, 434 N.E.2d at 775.

199. *Arriaga*, 163 F. Supp. 2d at 1192-93.

200. Arbitration clauses often deprive plaintiffs of the rights to a trial by jury and the opportunity to bring class action lawsuits. *See, e.g.*, Gipson v. Cross Country Bank, 294 F. Supp. 2d 1251, 1261-64 (M.D. Ala. 2003); Bank One, N.A. v. Coates, 125 F. Supp. 2d 819, 834 (S.D. Miss. 2001); Muhammad v. County Bank of Rehoboth Beach, 877 A.2d 340, 353-55 (N.J. Super. Ct. App. Div. 2005) (finding arbitration clauses deprive plaintiffs of the opportunity to bring class action suits).

201. The past several years have witnessed an intense scholarly focus on predatory lending in the mortgage context. *See* Tania Davenport, Note, *An American Nightmare: Predatory Lending in the Subprime Home Mortgage Industry*, 36 SUFFOLK U. L. REV. 531 (2003); *A Tale of Three Markets, supra* note 79; Elizabeth Renuart, *Toward One Competitive and Fair*

mortgage context, a definition of predatory lending has begun to take shape. Most experts in the area agree that it has some or all of the following characteristics: Predatory loans may be made at interest rates and may require fees that are higher than necessary to compensate for risk;[203] they may contain

---

Mortgage Market: Suggested Reforms in A Tale of Three Markets *Point in the Right Direction*, 82 TEX. L. REV. 421 (2003). Many articles have focused on the role and effect of legislation in the predatory lending debate. *See* Baher Azmy, *Squaring the Predatory Lending Circle*, 57 FLA. L. REV. 295 (2005); Kathleen C. Engel & Patricia A. McCoy, *The CRA Implications of Predatory Lending*, 29 FORDHAM URB. L J. 1571 (2002); Schiltz, *supra* note 7. Predatory lending has also been discussed in the context of small loans. *See* Creola Johnson, *Payday Loans: Shrewd Business or Predatory Lending?*, 87 MINN. L. REV. 1 (2002). Mandatory arbitration clauses in predatory lending contracts have also been the subject of scholarly attention. *See* Mark Berger, *Arbitration and Arbitrabililty: Toward an Expectation Model*, 56 BAYLOR L. REV. 753 (2004); Amy J. Schmitz, *Mobile-Home Mania? Protecting Procedurally Fair Arbitration in a Consumer Microcosm*, 20 OHIO ST. J. ON DISP. RESOL. 291 (2005); Robert Alexander Schwartz, *Can Arbitration Do More for Consumers? The TILA Class Action Reconsidered*, 78 N.Y.U. L. REV. 809 (2003). Many scholars and practitioners have also documented the negative social implications that may result from predatory lending. *See* Regina Austin, *Of Predatory Lending and the Democratization of Credit: Preserving the Social Safety Net of Informality in Small-Loan Transactions*, 53 AM. U. L. REV. 1217 (2004); Donna S. Harkness, *Predatory Lending Prevention Project: Prescribing a Cure for the Home Equity Loss Ailing the Elderly*, 10 B.U. PUB. INT. L.J. 1 (2000); Cecil J. Hunt, II, *In the Racial Crosshairs. Reconsidering Racially Targeted Predatory Lending Under a New Theory of Economic Hate Crime*, 35 U. TOL. L. REV. 211 (2003).

202. Borrowers who have been victims of credit card issuers, on the other hand, have sought legal action on predatory lending grounds. *See* Jennifer Harmon, *Broker Sues Credit Card Lenders*, NATIONAL MORTGAGE NEWS, Oct. 1, 2004, at 44, *available at* 2004 WLNR 189917 (The former President of the Florida Association of Mortgage Brokers sued nine lenders including Citicorp and Providian after he and his wife felt they had been victimized by being charged usurious rates on their credit cards. Part of the plaintiff's claim was designed to create a new cause of action in Florida called "Tortious Predatory Lending."). Consumer advocacy groups as well as certain lenders have been active in trying to prevent borrowers from being victims of credit card predatory lending. *See* Center For Responsible Lending, http://www.responsiblelending.org (last visited Sept. 13, 2006); Mortgage Bankers Association, http://www.stopmortgagefraud.com (last visited Sept. 13, 2006); National Association of Federal Credit Unions, http://www.nafcu.org (last visited Sept. 13, 2006); National Consumer Law Center, *Advance Notice of Proposed Rulemaking Review of the Open-End (Revolving) Credit Rules of Regulation* 2, http://www.nclc.org/initiatives/test_and_comm/content/open_end_final.pdf [hereinafter *Revolving Credit*]; National Consumer Law Center, http://www.consumerlaw.org (last visited Sept. 13, 2006).

203. *A Tale of Three Markets*, *supra* note 79, at 1265-66; Mansfield, *supra* note 20, at 535-47; Departments of the Treasury and Housing and Urban Development, *Curbing Predatory Home Mortgage Lending* 21 (June 20, 2000) http://www.huduser.org/publications/pdf/treasrpt.pdf [hereinafter HUD Report]. One study suggests that rate-risk disparities cost United States borrowers $2.9 billion annually and are fueled by the prevalence of incentives to brokers in the form of yield-spread premiums (YSP) to "steer" borrowers into loans that charge higher rates than are justifiable based on the risk posed

particularly onerous terms, such as large balloon payments and prepayment and late payment penalties.[204] They may be "structured to result in seriously disproportionate net harm to borrowers,"[205] or they may involve fraud, deception, or "lack of transparency."[206] They may "require borrowers to waive meaningful legal redress."[207] They may involve "loan flipping," whereby the lender makes a series of loans to the same borrower over a short period of time, charging new fees and sometimes higher interest rates each time.[208] Another overarching definition is that predatory loans are typically made to borrowers who cannot access the mainstream credit market either because of low income, racial minority status, or both.[209]

---

by the borrower. Eric Stein, Coalition for Responsible Lending, *Quantifying the Economic Cost of Predatory Lending* 9-10 (Oct. 30, 2001), http://www.responsiblelending.org/pdfs/Quant10-01.pdf.

204. Balloon payments require the borrower to pay off the principal due, or a large portion of it, in one payment at the end of the loan term. Mansfield, *supra* note 20, at 556. Balloon payments were incorporated into approximately 10% of all subprime loans in 1999; however, this figure varies by lender with at least one lender including balloon payments in 50% of its subprime loans. HUD Report, *supra* note 203, at 96. Prepayment penalties are designed to "reduce the risk of prepayment and compensate the lender for any cost resulting from prepayment." *Id.* at 93. There is a shocking disparity between the prevalence of prepayment penalties in the prime market (0.5% to 2% of all loans) versus the subprime market (70-76% of all loans). *Id.*

205. Engel & McCoy, *supra* note 201, at 1260.

206. *Id. See also* HUD Report, *supra* note 203, at 22 (describing unscrupulous tactics some lenders employ to hide costs from borrowers).

207. *A Tale of Three Markets*, *supra* note 79, at 1260.

208. Mansfield, *supra* note 20, 548-50. Another source describes "loan flipping" as follows:

> Loan flipping generally refers to repeated refinancing of a mortgage loan within a short period of time with little or no benefit to the borrower. Loan flipping typically occurs when borrower is unable to meet scheduled payments, or repeatedly consolidates other unsecured debts into a new, home-secured loan at the urging of a lender.

HUD Report, *supra* note 203, at 73.

209. Mansfield, *supra* note 20, at 559-61. According to one consumer advocacy organization, "Minority homebuyers are more likely to receive a subprime home purchase loan than white homebuyers. African-American homebuyers were 3.6 times more likely to receive a subprime home purchase loan than whites while Latinos were 2.5 times more likely to receive the subprime loan." ACORN Report, *supra* note 67, at 3 (emphasis omitted). While borrowers in the subprime and predatory markets are traditionally thought to be unable to access prime rate loans due to a negative financial history, there is evidence that a significant portion of subprime mortgage borrowers actually qualify for prime market rates. Freddie Mac reported that 10-35% of its borrowers were found to qualify for prime market rated loans. Freddie Mac, *Automated Underwriting Report* ch.5 (1996), http://www.freddiemac.com/corporate/reports/moseley/chap5.htm (citation omitted). Another study suggests that 50% of subprime borrowers qualified for prime market rates. *Id.* (citation omitted). The costs to these borrowers are significant. According to one study, a borrower in the subprime market will pay at least $200,000 more to a lender for a 30-year mortgage than a

This definition of predatory lending is consistent with the ignorant borrower paradigm in that it identifies specific ways in which lenders consistently exploit borrowers' lack of sophistication and information by imposing unfair and onerous terms in the bargaining process. Some predatory loan terms, such as high interest rates and high late payment penalties, had been regulated by state statute or case law prior to the deregulation of mortgage lending.[210] Case law actively policed other types of predatory behavior, such as loans causing net harm to borrowers or involving fraud or deception.[211] It is troubling that federal law has preempted some of these status-based protections, but they at least had the effect of defining certain market behavior as predatory.[212] In other words, these means of exploitation are considered aberrant and inappropriate in the mortgage market.

One of the fascinating things about mainstream credit card lending is that it has so many of the characteristics of predatory mortgage lending. Consider again the *Arriaga* case. Even ignoring the allegedly fraudulent scheme of pushing borrowers over their credit limits, that credit transaction involved high fees, a lack of transparency in that the lender regularly failed to send monthly account statements, and an arbitration clause which resulted in waiver of rights to a jury trial and a class action lawsuit as a condition of using the credit card.[213] These three characteristics, along with high interest rates,[214] are

---

similarly situated borrower in the prime market. ACORN Report, *supra*, at 67. One well-cited study estimates that predatory lending costs American borrowers $9.1 billion per year. Eric Stein, *supra* note 203, at 2.

210. *See supra* notes 107-109 and accompanying text; *see also* Saunders & Cohen, *supra* note 67, at 4 (providing historical background). North Carolina is one of the most active states in legislating to protect consumers from abusive lending practices, despite increased federal deregulation of the market. At least one study indicates this legislative action has decreased the number of predatory loans considerably. Comptroller of the Currency, *OCC Working Paper: Economic Issues in Predatory Lending* 2 (July 30, 2003), http://www.selegal.org/occ_workpaper0730.pdf [hereinafter OCC Working Paper].

211. Some common law claims still police the most egregious behavior. *See, e.g.*, Richter, S.A. v. Bank of Am. Nat'l Trust & Sav. Ass'n, 939 F.2d 1176 (5th Cir. 1991); Greene v. Gibraltor Mortgage Inv. Corp., 488 F. Supp. 177 (D. D.C. 1980); Peoples Trust & Sav. Bank v. Humphrey, 451 N.E.2d 1104 (Ind. Ct. App. 1983). In *Family Financial Services, Inc. v. Spencer*, 677 A.2d 479, 481, 485 (Conn. App. Ct. 1996), for example, the court held that the lender behaved unconscionably in making a second mortgage loan requiring a monthly payment of $733.33 to a borrower whose first mortgage monthly payments were $1011.00 and who had a monthly income of only $1126.67. Historical judicial contextualization of the mortgage borrower's story may add to the potency of the common law claims especially.

212. *See* Consumer Credit Protection Act, 15 U.S.C. §§ 1601-1693r (2000)

213. Arriaga, v. Cross Country Bank, 163 F. Supp. 2d 1189, 1190-91 (S.D. Cal. 2001); *see supra* Part II.A.ii.

214. After the introductory period during which a "teaser rate" is in effect, rates can increase by as much as 24%. Jeff Wuorio, Money Matters, *The Danger of Spreading Out Credit Card Debt* (July 15, 2002), http://www.bankrate.com/brm/news/money-matters/20020715a.asp. Often, one mistake can trigger the jump from the introductory rate to a

virtually universal in credit card lending.[215] The switch from low introductory rates to often much higher, permanent rates with little or no warning or accompanying information is a classic example of lack of transparency in the credit card market.[216]

Meanwhile, credit card borrower behavior remains unsophisticated. Extensive research shows that borrowers do not understand the implications of borrowing on a high interest credit card.[217] Research supported by innumerable

---

much higher permanent rate. For example, Chase's Platinum Visa Card reserves the right to increase the annual percentage rate (APR) to 29.99% for one late payment plus late fees, and the AT&T Universal Card increases the interest rate from zero to 28.49%. Lucy Lazarony, *Play Your Cards for an Interest-Free Loan* (June 23, 2005), http://www.bankrate.com/brm/news/cc/20030207a1.asp; Lucy Lazarony, *Step Slowly onto the Balance-Transfer Bandwagon* (Sept. 20, 2004), http://www.bankrate.com/brm/news/cc/20010312a.asp; *see also* John Rao, *Bankruptcy Bill's "Truth-in-Lending" Provisions Will Obscure the Truth* (Mar. 1, 2005), http://www.consumerlaw.org/initiatives/bankruptcy/tilprovisions.shtml (indicating that the new bankruptcy act allows creditors to easily alter the advertised teaser rates).

215. The OCC has recognized that "high interest rates and fees, balloon payments, high loan-to-value ratios, excessive prepayment penalties, loan flippings, loan steering and unnecessary credit insurance" occur in the credit card industry as well as the home mortgage industry. OCC Working Paper, *supra* note 210, at 6. Indeed, the National Consumer Law Center has described the entire market for credit-card credit as abusive. *See Revolving Credit*, *supra* note 202, at 2. The lack of transparency does not ubiquitously take the form of failure to send monthly statements but may involve changes in interest rates and other loan terms without adequate notice to the consumer. One classic manifestation is a universal default rate hike or assessment of penalty fee that is triggered by factors known only by the lender. For example, Consumer Action's 2005 Credit Card Survey found that the following circumstances can trigger a rate hike: decreased credit score (triggered a rate hike 90.48% of the time), paying another creditor late (85.71%), exceeding one's credit limit (57.14%), check bouncing (52.38%), carrying excessive debt (42.86%), having too much available credit (33.33%), obtaining another card (33.33%), and asking about mortgages or auto loans (23.81%). Consumer Action, *2005 Credit Card Survey*, CONSUMER ACTION NEWS 3 (2005), *available at* http://www.consumer-action.org. Negative amortization is also widespread in the industry. *See* Mara Der Hovanesian, *Tough Love for Debtors*, BUSINESS WEEK ONLINE (Apr. 25, 2005), http://www.businessweek.com/bwdaily/dnflash/apr2005/nf20050414_5876_db016.htm.

216. *See* Bar-Gill, *supra* note 5, at 1392-94. Lenders offer first-time consumers or consumers transferring debt from other cards low, sometimes 0%, introductory interest rates for a set number of months, at which point the introductory rate ends and a substantially higher interest rate kicks in. *Id.* The majority of charge activity occurs after the teaser rates have expired suggesting that most consumers are not vigilantly transferring balances to other cards with teaser rates or are unaware that the introductory offer has ended. *Id.* Typically, teaser rates only last for about 6 months. *Id.* Frequently, only the consumer who has a high credit rating will qualify for the advertised teaser rate. Bankrate.com, *Step-By-Step Guide to Balance Transfers* (Mar. 1, 2005), http://www.bankrate.com/brm/green/cc/basics4-3a.asp. Credit companies might promote a teaser rate of 3.9% that goes to 17%, but the consumer with less then perfect credit could be charged up to 7.9% initially and then 21% after the teaser expires. *Id.*

217. The Consumer Federation of America and Providian conducted a joint survey in 2004

anecdotes also shows that borrowers are surprised by the high fees imposed for exceeding credit limits, late payments and similar breaches of the credit card contract.[218] Generally, borrowers do not shop for better deals on terms such as these.[219] Even if they are sophisticated enough to understand the language in their contracts describing these terms, the terms are so standard in the industry that borrowers have neither the option to get credit elsewhere nor the power to bargain them away.[220]

The consequence is that the information asymmetry that is the hallmark of the predatory mortgage market permeates the mainstream credit card market. Credit card lenders understand the credit market intimately, and they deluge borrowers with advertisements for credit cards and related products that appear highly desirable.[221] Meanwhile, borrowers rarely read the fine print in the advertisements and contracts they receive, relying on the statements in bright,

---

that found that only 34% of respondents understood what a credit score indicates to a lender and only 12% knew that subprime rates are triggered with scores in the low 600's. Consumer Federation of American and Providian, *Most Consumers Do Not Understand Credit Scores According to a New Comprehensive Survey* (Sept. 21, 2004), http://www.consumerfed.org/pdfs/092104creditscores.pdf. Worse still, 28% of those surveyed said they believed spending the full amount of credit allotted on a card would improve their score. *Id.* Moreover, 34% of credit card borrowers have no idea what interest their lenders are charging. Nancy Ten Kate, *Debt Perception*, 19 AM. DEMOGRAPHICS 31 (1997).

218. One financial advisor reported that his clients are consistently surprised by fees in their statements including those generated when a credit card company changes the payment cycle dates without notice, alters interest rates dramatically, and promotes teaser rate schemes on balance transfer offers that ultimately cost the consumer heavily. Ray Martin, CBS NEWS, *More Credit Card Fees* (June 10, 2004), http://www.cbsnews.com/stories/2004/06/09/earlyshow/contributors/raymartin/main622103.shtml. Meanwhile, late fees and other penalties generated $1 billion dollars in 2004 for credit card companies. *Id.* "In 2000, fee income accounted for 25 percent of credit card companies' total income, and between 1995 and 1999, total fee income increased by 158 percent, from $8.3 billion to $21.4 billion." Bradley Dakake, *Deflate Your Rate: How to Lower Your Credit Card APR* 2 (March 27, 2002), http://truthaboutcredit.org/deflateyourrate3_02.pdf.

219. *See* Bar-Gill, *supra* note 5, at 1376-77 (arguing that consumers do not shop for better terms, but rather choose cards based on teaser rates and incentive offers because they do not believe when first applying for a card that they will incur a high debt or fail to make a payment on time).

220. *See* Anne Zeiger, *Lawmakers Take Aim at Credit Card Hikes* (Sept. 1, 2004), http://www.consumeraffairs.com/news04/credit_card_interest.html (Most major bank cards include the "universal default" policy in their contract terms.); *see also Revolving Credit*, *supra* note 202, at 26 (arguing that consumers do not have the choice "to avoid the onerous and abusive terms of open-end credit" because credit card issuers are "in a reverse competition to provide the most exploitative terms of credit" (emphasis omitted)); *see generally* Bar-Gill, *supra* note 5, at 1407 (Most card owners do not know that the onerous contract terms exist; therefore, they do not consider bargaining around the terms.).

221. Bar-Gill, *supra* note 5, at 1420-21; Laurie A. Lucas, *Integrative Social Contracts Theory: Ethical Implications of Marketing Credit Cards to U.S. College Students*, 38 AM. BUS. L.J. 413 (2001); *Revolving Credit*, *supra* note 202, at 18.

bold letters to make decisions about which card to obtain.[222] Compounding these asymmetries are the well-documented biases that seduce borrowers into believing that they will not borrow as much as they ultimately do.[223] This problem will only worsen as lenders market to more low-income borrowers who may believe that they have no other options for obtaining credit.[224]

Yet the behavior of credit card lenders is not characterized as predatory, aberrant, or inappropriate; instead, it is very well protected by credit card law. As *Arriaga* exemplifies, the contract rules supporting the enlightened borrower paradigm affirm loan terms generated by information asymmetry as an acceptable consequence of private bargaining.[225] Credit card case law assumes equal bargaining power and sophistication. As a consequence, credit card law has validated terms that mortgage lending historically found to be unacceptable.[226] While mortgage law's ignorant borrower paradigm delegitimizes certain lending activities as unacceptably exploitative, credit card law's enlightened borrower paradigm legitimizes these same lending activities as benefits of the bargain.

The new enlightened borrower paradigm has a broader legacy than the limitation on borrowers' rights and remedies described in the previous section. Its legacy is also to make a much broader range of lender behavior, including the exploitation of information asymmetries, positively normative. Indeed, it has made credit itself positively normative.[227]

C. *In Congress: The Response As Epitomized by Federal Legislation*

The law has not failed completely to recognize that abuses exist in the credit card market; however, the legislative response to abusive practices in the mortgage and credit card markets is entirely consistent with the dominant paradigms in each. Although similarly predatory practices exist in both

---

222. *Revolving Credit*, *supra* note 202, at 26.

223. *See, e.g.*, Lawrence M. Ausubel, *The Failure of Competition in the Credit Card Market*, 81 AM. ECON. REV. 50 (1991); Bar-Gill, *supra* note 5, at 1395-1401; Forrester, *supra* note 13, at 383-87 (discussing these biases in the context of home equity lending).

224. Credit card lenders market aggressively to many different demographics and are now increasingly targeting low-income borrowers. SULLIVAN ET AL., *supra* note 5, at 135-36; National Consumer Law Center, *Predatory Small Loans: A Form of Loansharking: The Problem, Legislative Strategies, A Model Act*, http://www.consumerlaw.org/initiatives/payday_loans/pay_menu.shtml ("From 1993 to 1996, the proportion of households with incomes under $20,000 who received credit card offers rose from 40% to 50%.").

225. Arriaga v. Cross Country Bank, 163 F. Supp. 2d 1189, 1194-95 (S.D. Cal. 2001) (determining that Arriaga was an enlightened borrower led the court to assume that she knew about and assented to the arbitration clause she later sought to have declared unconscionable).

226. *Compare Id.* (validating a credit card arbitration clause), *and* McGill v. Biggs, 434 N.E.2d 772 (Ill. App. Ct. 1982) (finding a mortgage instead of a warranty deed using extrinsic evidence).

227. Part III provides a more detailed discussion of this evolution in norms.

markets, Congress has enacted two markedly different statutes to counteract them; the statute regulating mortgages defines certain lending terms as abusive[228], while the statute regulating credit cards defines the failure to adequately disclose such terms as abusive.[229]

1. The Home Ownership and Equity Protection Act

The Home Ownership and Equity Protection Act of 1994 (HOEPA)[230] has been much maligned by consumer advocates.[231] It manages to eliminate only the most egregious abuses in the predatory lending market[232] and is very easy for lenders to evade.[233] At the same time, consumer advocates and others in the field regularly refer to HOEPA as a "de facto usury law," making it the only federal law capping interest rates on consumer loans in any respect.[234] And though it is less protective than many state laws, HOEPA's protections are significantly more interventionist than those of the disclosure regime. For example, HOEPA limits prepayment penalties,[235] bans negative amortization,[236] and prohibits balloon payments on short-term loans.[237] The statute effectively

---

228. *See* 15 U.S.C. § 1639(c), (e), (f).

229. See, e.g., Bankruptcy Abuse Prevention and Consumer Protection Act of 2005, 15 U.S.C. § 1637(c) (2005) (focusing on improving disclosure rather than on abusive actions of lenders).

230. 15 U.S.C. § 1601 (2000).

231. *See* Azmy, *supra* note 201, at 352-56; *A Tale of Three Markets*, *supra* note 79, at 1312; Debra Pogrund Stark, *Unmasking the Predatory Loan in Sheep's Clothing: A Legislative Proposal*, 21 HARV. BLACKLETTER L.J. 129, 144 (2005); *Protecting Homeowners: Preventing Abusive Lending While Preserving Access to Credit: Hearing Before the Subcomms. on Housing and Community Opportunity and Financial Institutions and Consumer Credit*, 10-11 (2003) (testimony of Margot Saunders), http://www.consumerlaw.org/initiatives/predatory_mortgage/content/AccessToCredit.pdf [hereinafter *Protecting Homeowners*].

232. Azmy, *supra* note 201, at 355-56.

233. Although recent changes have eliminated some loopholes, lenders can simply offer mortgage rates slightly below the statute's annual percentage rate (APR) floor or utilize financing options that HOEPA does not cover, such as purchase-money mortgages and open-ended credit lines. *See* Azmy, *supra* note 201, at 352-56; *A Tale of Three Markets*, *supra* note 79, at 1308.

234. *See, e.g.*, Christopher L. Peterson, *Federalism and Predatory Lending: Unmasking the Deregulatory Agenda*, 78 TEMP. L. REV. 1, 63 (2005).

235. 15 U.S.C. § 1639(c). The term prepayment penalty is liberally applied to include any method of accruing unearned interest that places the consumer in a less favorable position than the actuarial method. *Id.* Prepayment penalties are allowed, however, if the consumer is refinancing through a different creditor than that of the original mortgage, or if the consumer's mortgage payment is less than a specified percentage of her income as verified through fiscal records. 15 U.S.C. § 1639(c)(2).

236. 15 U.S.C. § 1639(f). The consumer's mortgage payments must compensate fully for the interest due, thus decreasing the balance with each payment. *Id.*

237. 15 U.S.C. § 1639(e). For mortgages lasting less than five years, the consumer must be

caps interest rates, mainly by expanding liability to assignees of above-market interest loans.[238] The statute even provides a very narrow version of the suitability requirement proposed by Professors Engel and McCoy as a way to curb predatory lending.[239] In these respects, HOEPA restores some of the state law substantive protections that were eliminated by AMTPA and DIDMCA.

It is not surprising that substantive protection arises in the realm of mortgage lending. Consistent with the ignorant borrower paradigm, HOEPA addresses information asymmetries in the predatory mortgage market by overriding particularly egregious terms of the lending contract.[240] Rather than requiring more complete disclosure to the borrower about the consequences of accepting such terms, the statute simply prohibits certain types of penalties and payment terms.[241]

2. The Bankruptcy Abuse Prevention and Consumer Protection Act of 2005

In credit card lending, the recent legislative attempts to curb abuse are best represented by the Bankruptcy Abuse Prevention and Consumer Protection Act of 2005 (BAPCPA), which has received much attention for its sweeping restrictions on consumer bankruptcy.[242] A less well-publicized part of the Act, however, seeks to protect credit card borrowers by requiring more stringent disclosure.[243]

The statute requires that lenders on open-credit accounts (which include virtually all credit card accounts) that require a minimum monthly payment of less than a certain percentage of the overall debt owed must include prominent text on the front of the billing statement that reads as follows:

> Minimum Payment Warning: Making only the minimum payment will increase the interest you pay and the time it takes to repay your balance. For example, making only the typical 2% minimum monthly payment on a

---

able to amortize the balance through the scheduled payment terms and lenders cannot require additional payments that increase the loan balance. *Id.*

238. 15 U.S.C. § 1641(d)(1) (2005); *see generally* Kurt Eggert, *Held Up In Due Course: Predatory Lending, Securitization, and the Holder in Due Course Doctrine,* 35 CREIGHTON L. REV. 503 (2002) (discussing the holder-in-due-course doctrine and its impact on predatory lending).

239. HOEPA's suitability requirement prohibits creditors from engaging in a "pattern or practice" of awarding mortgages to consumers without considering their repayment aptitude in light of their financial status. 15 U.S.C. § 1639(h). Professors Engel and McCoy conclude that HOEPA's provision is too narrow to be effective and call for congressionally mandated self-regulation, an industry power to regulate, and a federal cause of action for breaching the standard's imposed duties. *A Tale of Three Markets*, *supra* note 79, at 1337-44.

240. *See* 15 U.S.C. § 1639(e).

241. *See Id.*

242. Bankruptcy Abuse Prevention and Consumer Protection Act of 2005, 15 U.S.C. § 1637 (2005).

243. *Id.*

> balance of $1,000 at an interest rate of 17% would take 88 months to repay the balance in full. For an estimate of the time it would take to repay your balance, making only minimum payments, call this toll-free number: _____ (the blank space to be filled in by the creditor).[244]

The statute also requires prominent warnings regarding the timing and amounts of late payment penalties that may be charged,[245] as well as the terms of any introductory or teaser interest rates that may be used.[246]

The federal statutory response to the credit card market continues to focus on improving disclosure.[247] Rather than capping late payment penalties to the amount required to cover the lender's costs and risks, or banning them outright, the statute requires more prominent warnings.[248] In doing so, the most recent law on credit cards demonstrates a continuing affinity for the enlightened borrower paradigm. Credit card law does not define exorbitant late payment penalties and negative amortization as inherently abusive, but rather it defines the practice of withholding information about such terms as abusive. Congress has taken this approach despite evidence that the information asymmetries in the mainstream credit card market are as extreme, if not more so, than in the predatory mortgage market.

III. WHAT DOES THE FUTURE HOLD?

As Americans continue to borrow more, consume more, and save less, the likelihood of continuing abuses in the consumer credit market appears high. One of the critical questions for lawmakers in the next decade will be whether the disclosure regime of the last four decades is the appropriate choice for balancing access to credit against exploitation of the consumer borrower. The response to this question should be a conscious decision rather than a default rule based on either tradition or convenience. Part III.A. predicts the future of consumer credit regulation if lawmakers stay their current course. Part III.B. raises some issues that lawmakers should consider in regulating the credit card market.

A. The Current Course of Regulation

It is possible that the ignorant and enlightened borrower paradigms will exist in equipoise in the two major markets for consumer credit, but it does not seem likely. Recent developments in the residential mortgage market signify an increasing shift towards the disclosure regime of the credit card world. The

---

244. Bankruptcy Abuse Prevention and Consumer Protection Act of 2005, 15 U.S.C. § 1637(b) (2005).

245. 15 U.S.C. § 1637(b) (2005).

246. 15 U.S.C. § 1637(c) (2005).

247. *Id.*

248. 15 U.S.C. § 1637(b) (2005).

deregulatory trend continues in the banking industry, preempting many state consumer protection laws.[249] The Ney-Kanjorski bill, if enacted, could result in much broader preemption of substantive state law protections for the borrower.[250] Meanwhile, Congress has focused much of its consumer protective energy on tweaking RESPA[251] and HOEPA[252] to impose better disclosure requirements on mortgage lenders. Although HOEPA does impose substantive protections for the mortgage borrower, it is not nearly as effective at curbing abuses as its state law predecessors and successors. Overall, mortgage law is becoming more contractarian and less status-based. If this trend continues, the ignorant borrower paradigm could eventually be a legal imprint of the past.

Thus far, consumer advocates have opposed these developments, arguing instead for more substantive consumer protections, such as usury caps, and elimination of the exportation clause and other instruments of preemption, which limit the applicability of more protective state laws.[253] These efforts are well-intentioned, admirable, and, in some respects, effective.[254] Such efforts may lead to more HOEPA-like protections. Abuses in the credit card industry may garner enough attention to generate an open-end credit statute that caps certain types of penalties, prohibits negative amortization, and imposes usury caps. Perhaps the BAPCPA is a first step in that direction. Lawmakers may also scrutinize more thoroughly abuses that are typically outside the realm of private ordering, such as advertising, which occurs before a contract is formed, and debt collection practices, which are not easy to anticipate at the time of contract formation.

These efforts, however, often do not attack disclosure for being largely ineffective, thus implicitly affirming lawmakers' predilection to regulate disclosure only.[255] Academic literature has detailed the failures of the credit

---

249. *See, e.g.*, Wachovia Bank, N.A. v. Burke, 414 F.3d 305, 309 (2d Cir. 2005) (holding that the National Bank Act preempts Connecticut laws subjecting national bank subsidiaries to licensing and other requirements).

250. *See supra* note 77.

251. *See* THE COST OF CREDIT, *supra* note 61, Supp. 2004 at 253, 256-58.

252. *See* Judith M. Scheiderer, *TILA Regulation and Litigation Update*, 59 CONSUMER FIN. L. Q. REP. 162, 163-64 (2005); THE COST OF CREDIT, *supra* note 61, Supp. 2004 at 284.

253. *See, e.g.*, S.P. Dinnen, *Group Targets High Credit Card Fees*, DES MOINES REGISTER (Nov. 2, 2003), *available at* http://www.demos-usa.org/pubs/Group_Targets_High_Credit_Fees.pdf (arguing for a usury cap on credit card loans); *Protecting Homeowners*, *supra* note 231, at 11-13 (advocating for elimination of fees and points on certain types of mortgage loans); U.S. PIRG Reports, *Show Me The Money: A Survey Of Payday Lenders And Review Of Payday Lender Lobbying In State Legislatures* (Feb. 1, 2000), http://uspirg.org/resports/paydayloans2000/showmethemoneyfinal.pdf (arguing for a federal usury cap and for limitations on preemption).

254. *See, e.g.*, Zieger, *supra* note 220(describing a proposed New York statute that would prohibit credit card issuers from raising interest rates or fees on the basis of unrelated activity on a cardholder's credit report).

255. *But see Revolving Credit*, *supra* note 202, at 25-27 (arguing that abuses in the credit

card disclosure requirements that currently exist.[256] Although some of these authors have proposed improved forms of disclosure,[257] it is hard to imagine that these new reforms will work. Disclosure fails for too broad a range of reasons to be rehabilitated. For example, although the BAPCPA requires new warnings stating the number of months required to repay a loan, [258] it does not require a potentially more attention-grabbing statement of the total amount owed at the end of that time period.[259] Nor is there any evidence that consumers will be alarmed by merely a statement of a time period required to repay a loan. Even if a statement of the total amount owed were statutorily required, it is hard to believe it could make much impact when consumers so rarely read their credit card contracts. Borrower education proposals face some of the same difficulties.[260] Consumers are already too broadly acculturated to ignoring fine print, to the ease of applying for credit cards, and to the overall lack of gravity in credit card borrowing.

Consequently, the incremental changes under way, many of which are still anchored in the disclosure regime, will be ineffective in the face of increasingly aggressive practices of lenders in both markets. This Article does not propose that mortgage law, either in its current or historical form, is the answer to the problems in the credit card or predatory mortgage lending markets. As the continuing abuses in the subprime mortgage market demonstrate, the combined effect of rights of redemption, foreclosure protections, and judicial reinterpretations of the loan agreement are insufficient to stop predatory lending. It is questionable whether more stringent usury caps would even suffice.

Nor does this Article assert that the distinction between status and contract is either wholly accurate or ultimately even meaningful. After all, by seeking an expansive freedom to contract, credit card lenders seek legal enforcement of their more privileged status vis-à-vis their borrowers. This Article proposes, however, that mortgage law embodies a regulatory approach, grounded in protecting the particular status of an ignorant borrower, which is not evident in credit card law. While the individual substantive protections that have arisen from this status-based regulatory approach in mortgage law could be improved,

---

card market are too rampant to be solved by improved disclosures).

256. *See, e.g.*, Azmy, *supra* note 201, at 345-52; Bar-Gill, *supra* note 5, at 1417-20; Michael S. Barr, *Credit Where It Counts: The Community Reinvestment Act and Its Critics*, 80 N.Y.U. L. REV. 513, 628-33 (2005); Peterson, *supra* note 7, at 886.

257. *See, e.g.*, Bar-Gill, *supra* note 5, at 1419-20; Peterson, *supra* note 7, at 886-903.

258. *See supra* note 244 and accompanying text.

259. *Id.*

260. *See* Alan Schwartz & Louis L. Wilde, *Intervening in Markets on the Basis of Imperfect Information: A Legal and Economic Analysis*, 127 U. PA. L. REV. 630, 633-35 (1979); Stark, *supra* note 231, at 144-45; Abdighani Hirad & Peter M. Zorn, *A Little Knowledge Is a Good Thing: Empirical Evidence of the Effectiveness of Pre-Purchase Homeownership Counseling* 1-2 (2001), http:// www.jchs.harvard.edu/publications/homeownership/liho01-4.pdf.

the underlying approach has led to the greater level of substantive protections in mortgage law.

Could and should such a status-based approach also improve protections for credit card borrowers? Such protections might be justified if, as appears to be the case, the disclosure regime sustains an abusive market. Such protections would seem especially appropriate if credit card loans were accomplishing some of the same socially beneficial purposes as mortgage loans. Thus, integral to a decision concerning whether the disclosure regime is the right choice for the future is a deeper understanding of why the paradigm shift in consumer lending occurred and whether it is based on real differences between mortgage and credit card borrowing. The concluding section of this Article explores these questions.

B. Distinctions Without a Difference

Although the switch in borrower paradigms may seem both natural and rational, it is increasingly grounded in a set of distinctions without a difference. This section examines two explanations for the switch in paradigms and the assumptions underlying these explanations. In concluding that these assumptions are no longer true, this Article suggests that credit cards facilitate a new role for credit that serves purposes previously only served by secured credit. This new role for credit requires a new inquiry into the legitimacy of the enlightened borrower paradigm.

Two explanations appear to justify the enlightened borrower paradigm. Both have been the subject of too little scrutiny. The first is the lack of security in credit card lending. The security at issue in mortgage lending is typically the borrower's most unique and cherished asset, and it is likely that many modern-day judges and legislators can empathize with the catastrophic losses associated with this form of lending. The lender receives an extraordinary benefit from the threat to the borrower of loss of her home. Historically, the law has counterbalanced this benefit by preventing lenders from overreaching in their right to the security.[261] For example, early in mortgage law's history, courts enunciated the rule that the mortgagee's title did not signify ownership, but rather merely security.[262] Compounding this empathetic response is the historical maxim to promote utilization and alienation of land.[263] In the context of mortgage lending, this principle bodes in favor of keeping the land in the hands of the owner who is making active use of it, rather than the lender who merely has a security interest in the land. The former, by virtue of her past

---

261. *See* OSBORNE, *supra* note 19, at 13, 25.

262. *Id.*

263. *See, e.g.*, JOSEPH W. SINGER, PROPERTY LAW: RULES, POLICIES AND PRACTICES 557-58 (3d ed. 2002) (discussing the connection between this maxim and the modern law of estates); Oliver S. Rundell, *The Suspension of the Absolute Power of Alienation*, 19 MICH. L. REV. 235 (1921) (discussing the connection between this maxim and the promotion of land's alienability). This maxim also helps justify specific performance as a remedy in property cases.

behavior, has proven her interest in using the land, while the latter is untested in that regard.[264]

Credit card lenders who do not have the benefit of such potent security, might justifiably expect fewer legal constraints on their ability to protect themselves from the risks of unsecured lending. They might argue that the more onerous terms of their lending contracts reduce their risk in some of the same ways that collateral limits risk for a secured lender.[265] Onerous penalty clauses might deter borrowers from defaulting in the same way as does the threat of losing their homes.[266] Without such clauses, unsecured creditors may be unwilling to risk lending without any security.

The second explanation for the switch to the enlightened borrower paradigm is the existence of different roles for mortgages and credit cards in wealth formation. At the most basic level, the mortgage transaction connotes ownership because the quintessential use of a residential mortgage loan is to purchase the property used as collateral.[267] Every time a mortgagor makes a mortgage payment, she is building her equity.[268] In this respect, mortgage loans contribute directly to capital formation.

Moreover, by protecting the borrower's collateral, mortgage law has traditionally encouraged consumer participation in the market for land, both as a buyer via the mortgage loan and as a seller by sale of the home and use of the equity, ideally to purchase a more expensive home. Protection of the borrower's collateral perpetuates a market in residential real estate, while onerous lender's remedies, such as strict foreclosure, might chill such a market.[269] Such a chilling effect would not be limited to the borrower at issue. Consider, for example, the negative externalities resulting from loss of a home

---

264. The historical image of the home as one's castle, as described by Professor Singer in his recent analysis of takings law, probably contributes to judicial sensitivity to the plight of mortgage borrowers. *See* Joseph William Singer, *The Ownership Society and Takings of Property: Castles, Investments, and Just Obligations*, 30 HARV. ENVTL. L. REV. 309, 314-15 (2006). I am grateful to Professor Singer for pointing out that another context in which consumers are treated as enlightened is landlord/tenant law, where "enlightened" tenants historically have received less protection than "ignorant" homeowners.

265. A similar argument has been made to justify negative pledge and other covenants in unsecured lending to businesses. *See* Lucian Arye Bebchuk & Jesse M. Fried, *The Uneasy Case for the Priority of Secured Claims in Bankruptcy*, 105 YALE L.J. 857, 878-80 (1996); Carl S. Bjerre, *Secured Transactions Inside Out: Negative Pledge Covenants, Property and Perfection*, 84 CORNELL L. REV. 305, 306-11 (1999).

266. Moreover, the argument would continue, the backstop to such deterrence measures should be whether they are unconscionable, not whether they risk loss of a cherished asset because there is no obvious cherished asset at risk.

267. Home equity loans complexify this picture. Their relatively more recent development as a mainstay of consumer borrowing blurs the line between mortgage and credit card credit.

268. An exception here is financing with an interest-only component, for the period during which only interest is paid.

269. On the harshness of strict foreclosure, see OSBORNE, *supra* note 19, at 651-53; Wechsler, *supra* note 41, at 858-60.

to foreclosure. Foreclosure may result in a below-market sales price or, worse yet, in an empty, boarded-up home accompanied by the neighborhood deterioration that the broken windows theory would predict.[270] Mortgagor protections promote alienability, a fundamental goal in American property law. In these respects, the ownership society created by mortgage-based credit produces far-reaching social benefits.[271]

Credit cards do not so obviously contribute to wealth formation. Credit card borrowers do not use their loans to purchase land, nor do they build equity when they repay their loans. As described by Professors Sullivan, Warren, and Westbrook, credit card funds typically are used to purchase consumable goods and services, such as restaurant meals, clothing, and vacations.[272] Thus, while the quintessential role of mortgage-based credit is to facilitate ownership of an asset that generally appreciates in value, the quintessential role of credit card credit is to facilitate ownership of assets that depreciate quickly. Normally, such assets are not unique, and, thus, do not warrant specific performance remedies in the event of their confiscation.

Credit card credit produces market participation by borrowers as buyers but far less robust participation as sellers. The market for used clothes and sneakers typically has resale prices far below the original purchase prices, while there simply is no market for half-eaten pizza or unused fitness club memberships. In other words, while credit card credit is integral to keeping consumption high, it has a less obvious socially beneficial role in wealth formation.

If accurate, these two explanations may justify a less protective disclosure regime in credit card law, but they fail to explain why the choice for consumers between mortgage and credit card credit is so subtle. This is a troubling disconnect, and it demands careful research into the distinctions between secured and unsecured consumer credit. What is needed is an examination of how credit cards change the traditional assumptions that underlie these deep-rooted explanations for the switch in paradigms. In particular, two assumptions appear far more questionable since the ascendance of the credit card.

---

270. James Q. Wilson & George L. Kelling, *Broken Windows* 2-3 (1982), http://www.manhattan-institute.org/pdf/_atlantic_monthly-broken_windows.pdf. Meanwhile, it is commonplace that home ownership promotes pride and investment in one's property and more broadly in the community.

271. In addition to the rise of home equity loans, an important cross-current in the real estate world is the refinancing craze of the last several years. This phenomenon challenges the traditional assumptions about the benefits of mortgage lending by, for example, artificially inflating home values, encouraging the use of money from refinancing to purchase consumable goods rather than to save or invest and churning debt rather than creating new opportunities for home ownership. For statistics and comments about the refinancing phenomenon, see, e.g., William R. Emmons, *Consumer-Finance Myths and Other Obstacles to Financial Literacy*, 24 ST. LOUIS U. PUB. L. REV. 335 (2005); Freddie Mac, *Cash-Out Refinance Activity Rises in Second Quarter* (2005), http://www.freddiemac.com/news/archives/rates/2005/2qupb05.html.

272. SULLIVAN ET AL., *supra* note 5, at 111.

The first assumption is that physical[273] assets are necessary for credit to accomplish wealth formation. Said another way, conventional wisdom is that equity cannot be accumulated unless the loan is used to buy an asset that can later be resold. Without the ability to build equity in an asset, credit cannot serve the purpose of enhancing the wealth of the borrower. This assumption is built on the equally traditional beliefs that wealth consists of assets not debts and that it is socially beneficial to promote the attainment of such assets. Credit cards challenge this assumption. Using credit to build equity in a house is one important means to accomplish wealth formation, but it is no longer the only means. In our modern times, credit itself is an important form of wealth. For example, by enabling the purchase of designer clothing and fine dining, credit cards enhance social status,[274] as do other forms of wealth. Credit card credit also provides a means to acquire greater wealth and to invest in capital assets,[275] as do other forms of wealth. Moreover, the more an individual uses credit cards, the more credit she typically acquires by means of increased credit lines.[276] In this respect, too, credit card credit functions like other forms of wealth.

Notably, this use of credit cards is a new role for credit in American society, at least on such a large scale.[277] Prior to the ascendance of the credit card,[278] consumers rarely purchased consumable goods or services with credit, with the exception of the automobile loan.[279] To some, this is a positive development.[280] Credit cards give individuals more options to borrow. For those who do not own a home, credit cards open up a new world of market participation. While these new status enhancers appear to be gaining

---

273. I focus on physical collateral here because the home is the paradigmatic collateral.

274. Here, the term "status" signifies something different from a borrower's position in a class of borrowers, such as mortgagors, but rather connotes a rise in economic welfare, social welfare, or both.

275. *See, e.g.*, Creola Johnson, *Maxed Out College Students: A Call to Limit Credit Card Solicitations on College Campuses*, 8 N.Y.U. J. LEGIS. & PUB. POL'Y 191, 221 n.129 (2005) (describing a study that found on one campus "68.8% of students surveyed used their credit cards to pay for tuition, books, and supplies"). Many small business owners also use credit cards to start up or keep their businesses afloat. *See* SULLIVAN ET AL., *supra* note 5, at 115-19.

276. Credit Card Manual, *7 Simple Ways to Increase Your Credit Card Limit*, http://www.creditcardmanual.com/creditcard/7_Simple_Ways_to_Increase_Your_Credit_Card_Limit.html (last visited Sept. 14, 2006).

277. Certainly there have always been informal means of financing the purchase of consumer goods, *see, e.g.*, Austin, *supra* note 201, but both the formality and the breadth of such financing were introduced by the credit card.

278. This refers to the merchant credit cards of the early 20th Century. *See* SULLIVAN ET AL., *supra* note 5, at 109.

279. Note also that there is a much more robust market in used cars than in used clothes.

280. *See, e.g.*, Zywicki, *supra* note 6. *See also* Peter V. Letsou, *The Political Economy of Consumer Credit Regulation*, 44 EMORY L.J. 587 (1995).

importance in modern society, the more basic point here is that in this new role, credit cards accomplish wealth formation.[281]

The second questionable assumption is that something tangible must serve as collateral for the borrower to perceive enough risk that she will have an incentive to repay her loan. This is one of the most basic assumptions of secured lending.[282] While unsecured lenders compensate for this lack of security with higher interest rates and more onerous penalty clauses, conventional wisdom is that these deterrents can never be as effective as the risk of loss of one's home. The latter is catastrophic to borrowers across the board, while the former deters those who are relatively confident of repaying their loans, those who have assets that may be liquidated through judgment liens, or those who expect to borrow repeatedly from the same lender. Onerous terms do little to deter those one-time borrowers whose prospects are very risky and who have little to lose from default—or so the assumption goes.[283]

Credit cards challenge this assumption as well. Credit card borrowers encounter enormous risk in the borrowing process that is not limited to onerous loan terms and higher interest rates. Credit card borrowers risk their homes, physical assets, and much more. Because credit bureaus track their payment histories, borrowers risk their reputations as creditworthy individuals. If their credit ratings are negative, they risk loss of opportunities to obtain designer clothes, health club memberships, and rental cars. They also risk loss of credit. In this respect, they risk loss of significant wealth. It does not seem an overstatement to say that a person cannot be a member of the middle class in the United States without having an acceptable credit rating.

Credit cards have normalized the use of a new, non-physical form of collateral in consumer lending. Whether this new collateral is called reputation, honor, dignity, or shame, it is a powerful deterrent to default in the market for consumer credit, and it operates much like the threat of loss of one's home. Most of us, like Andrea Arriaga, cannot risk negative credit reports. In her case, this threat precluded her from opting out of the onerous contract imposed by the credit card lender.[284] Thus, credit card lenders do, in fact, receive an

---

281. This function is not necessarily through resale, as with houses.

282. *See* LYNN M. LOPUCKI & ELIZABETH WARREN, SECURED CREDIT: A SYSTEMS APPROACH 43 (Little, Brown & Co. 1995); David Gray Carlson, *On the Efficiency of Secured Lending*, 80 VA. L. REV. 2179, 2188-89 (1994); Ronald J. Mann, *Explaining the Pattern of Secured Credit*, 110 HARV. L. REV. 625, 645-48 (1997).

283. This reasoning leads banks to ration credit to more marginal borrowers. For discussions of this phenomenon, see Keith N. Hylton, *Banks and Inner Cities: Market and Regulatory Obstacles to Development Lending*, 17 YALE J. ON REG. 197, 207-17 (2000); Michael Klausner, *Market Failure and Community Investment: A Market-Oriented Alternative to the Community Reinvestment Act*, 143 U. PA. L. REV. 1561, 1565-66 (1995); *see e.g.*, Joseph E. Stiglitz & Andrew Weiss, *Credit Rationing in Markets with Imperfect Information*, 71 AM. ECON. REV. 393 (1981).

284. Arriaga v. Cross Country Bank, 163 F. Supp. 2d 1189 (S.D. Cal. 2001); *see supra* notes 168-70 and accompanying text.

extraordinary benefit in their transactions with consumers, one that mimics some of the important advantages of collateral.

In both these respects, the use of credit as wealth and the collateralization of dignity, the line between secured and unsecured borrowing is becoming blurred and the justifications for two different paradigms much less obvious for the consumer borrower.[285] While the law encourages mortgage borrowers' wealth-improving use of credit by protecting their ownership of the collateral, it fails to recognize the wealth formation accomplished by credit card borrowing and leaves these borrowers to suffer extraordinary losses with little legal protection. Moreover, this new role for credit suggests that both the current enlightened borrower legal regime and many of the reforms proposed by consumer advocates may be off the mark. Rather than drawing lines on the traditional battlefield between access and exploitation, the more productive strategy for lawmakers and advocates may be to consider how to enhance the wealth-forming potential of credit card borrowing.

Mortgage law's focus on the ignorant borrower provides a valuable place to begin redefining credit card policy. By reintroducing the borrower's status as a basis for protection, this Article intends to refocus the debate about credit cards on questions about their ability to fulfill the role for credit that they have helped to create. Among the important questions that ought to be explored is the extent to which the freedom of contract sought by credit card lenders imposes both unacceptable risks on credit card borrowers and negative externalities on those who have to suffer the social and economic consequences of catastrophic problems that can result from unregulated credit card use. What exactly are credit card borrowers risking? What is the nature of the losses that defaulting credit card borrowers suffer? How do these losses affect credit card borrowers' ability to enhance their wealth? And how can the law facilitate their recovery from loss and their continued participation in wealth formation? Questions such as these ought to elucidate the winners and the losers, those in greater need of protection, and those who require more regulation in the current market for credit card credit.

IV. CONCLUSION

In the market for consumer credit, might status, and not contract, be the way to progress? The emergence of the enlightened borrower paradigm in the law of credit card lending requires us to consider this question. Given the unfavorable legal and market consequences of switching to this paradigm, the answer may be yes.

285. Certainly, the law does not blur this line, but rather consumer behavior treats the two kinds of debt more interchangeably. Both kinds of debt do contribute to wealth formation.

PROCEDURAL DUE PROCESS TO DETERMINE "ENEMY COMBATANT" STATUS IN THE WAR ON TERRORISM

TUNG YIN*

How does the United States know that persons detained and classified as "terrorists" are who they appear to be? After the terrorist attacks on September 11, 2001, Congress authorized the President to use military force against "those nations, organizations, or persons [the President] determines planned,

* Associate Professor, University of Iowa College of Law. J.D., 1995, University of California, Berkeley (Boalt Hall). Thanks to Randy Bezanson, Stephanos Bibas, Arthur Bonfield, Bill Buss, Jon Carlson, Judge John Ferren, Jill Gaulding, Jonathan Hafetz, Kevin Johnson, Stephen Legomsky, Amy Liu, Sir Geoffrey Palmer, Katie Porter, Margaret Raymond, John Reitz, Wendie Schneider, Alexander Somek, and Ethan Stone for helpful discussions and feedback. Special thanks also to Robin Dull ('07) and Steven Stickle ('07) for research assistance.

authorized, committed, or aided the terrorist attacks that occurred on September 11, 2001, or harbored such organizations or persons."[2] President Bush determined that the terrorist group Al Qaeda, based in Afghanistan and sheltered by that country's leaders, the Taliban, was responsible for the 9/11 attacks.[3] President Bush authorized military action against both Al Qaeda and the Taliban, resulting in the capture of hundreds of persons—including two American citizens—and their eventual classification as "enemy combatants."[4] Nearly five years after the United States started dropping bombs in Afghanistan, some 455 suspected "enemy combatants" remain in detention at the U.S. naval base in Guantanamo Bay, Cuba.[5]

The executive branch has justified the continuing detention of suspected "enemy combatants" by reference to the law of armed conflict.[6] In a traditional war between nations, however, the members of armed forces fight in distinctive uniforms, carry their arms openly, and comply with the laws of war.[7] The result is that one's combatant status should be fairly apparent. But this is not so in the current war on terrorism. Al Qaeda members masquerade as civilians and refuse to wear distinctive uniforms or display other identifying marks.[8] Based on their noncompliance with the laws of war, the executive branch has asserted that captured detainees lack prisoner of war status.[9]

---

2. Authorization for Use of Military Force, Pub. Law No. 107-40, §2(a), 115 Stat. 224, 224 (2001). *See generally* Curtis A. Bradley & Jack L. Goldsmith, *Congressional Authorization and the War on Terrorism*, 118 HARV. L. REV. 2047 (2005) (providing a comprehensive analysis of the joint resolution authorizing the use of military force in 2001).

3. Jeanne Cummings & David Rogers, *Bush Says "Freedom and Fear are at War,"* WALL ST. J., Sept. 21, 2001, at A3.

4. Adam Liptak et al., *After Sept. 11, a Legal Battle on the Limits of Civil Liberty*, N.Y. TIMES, Aug. 4, 2002, § 1, at 1.

5. News Release, U.S. Dep't of Defense, Office of the Assistant Sec'y of Affairs (Sept. 14, 2006), *available at* http://www.defenselink.mil/Releases/Release.aspx?ReleaseID=9960 [hereinafter DOD News Release].

6. *See, e.g.*, Neil A. Lewis, *Administration's Position Shifts on Plans for Tribunals*, N.Y. TIMES, Nov. 2, 2002, at A8 ("[G]overnment officials are now confident that they may detain the Taliban and Qaeda prisoners indefinitely at Guantanamo Bay in Cuba and elsewhere as enemy combatants.").

7. *See* Geneva Convention Relative to the Protection of Prisoners of War art. 4, Aug. 12, 1949, 6 U.S.T. 3316, 75 U.N.T.S. 135 [hereinafter Geneva Convention].

8. *See, e.g.*, John C. Yoo & James C. Ho, *The Status of Terrorists*, 44 VA. J. INT'L L. 207, 217 (2003).

9. *See* Alberto R. Gonzales, Editorial, *The Rule of Law and the Rules of War*, N.Y. TIMES, May 15, 2004, at A17 (setting forth Attorney General's argument as to Al Qaeda and Taliban members' lack of entitlement to POW status); Katharine Q. Seelye, *Criticized, U.S. Brings Visitors to Prison Camp*, N.Y. TIMES, Jan 26, 2002, at A8 (discussing Al Qaeda members' lack of entitlement to POW status); Thom Shanker & Katharine Q. Seelye, *Word for Word/The Geneva Conventions; Who is a Prisoner of War? You Could Look It Up. Maybe*, N.Y. TIMES, Mar. 10, 2002, § 4, at 9 (quoting senior government official as asserting that the Taliban fail to qualify for POW status).

The very factor element that leads the United States government to argue that detainees have forfeited POW status—failure to distinguish oneself from noncombatants—also, however, substantially increases the risk that persons are detained erroneously, especially because the United States paid Northern Alliance fighters cash bounties for each captured Taliban or Al Qaeda fighter.[10] The increased likelihood of erroneous detention underscores the importance of the procedures used to classify detainees as enemy combatants, and therefore, the importance of determining whether the detainees have due process rights.

In the Supreme Court's October 2003 term, the Court granted *certiorari* on three cases resulting from the war on terrorism: (1) *Rasul v. Bush*[11]; (2) *Hamdi v. Rumsfeld*[12]; and (3) *Rumsfeld v. Padilla*.[13] In these three cases, the Court appeared ready to answer what procedures were constitutionally required to minimize erroneous classifications as enemy combatants. These three cases spanned the universe of relevant fact-patterns: *Rasul* and *Hamdi* involved alien and citizen detainees, respectively, captured on the battlefield,[14] while *Padilla* involved a citizen arrested within the United States.[15] The factors of citizenship and geography were significant because under existing precedent, it is undisputed that persons (alien and citizen alike) *within* the United States are entitled to due process and that citizens within or outside the United States are entitled to due process.[16] The uncertainty, however, lay in whether nonresident aliens outside the United States were entitled to due process.[17] When the Court issued decisions in all three cases on June 28, 2004,[18] it failed to resolve that uncertainty definitively.

---

10. Bill Dedman, *U.S. To Hold Most Detainees at Guantanamo Indefinitely*, BOSTON GLOBE, Apr. 25, 2004, at A1.

11. 542 U.S. 466 (2004).

12. 542 U.S. 507 (2004) (plurality decision).

13. 542 U.S. 426 (2004).

14. *See* 542 U.S. at 470-71; 542 U.S. at 510 (plurality decision).

15. *See* 542 U.S. at 430.

16. *See* Plyler v. Doe, 457 U.S. 202, 210 (1982) (holding aliens inside the United States entitled to equal protection); Burns v. Wilson, 346 U.S. 137 (1953) (holding U.S. citizens outside the United States entitled to habeas relief).

17. This Article uses the term "outside the United States" as a term of art; nonresident aliens who are deemed outside the United States *legally* are usually *physically* within the United States, i.e., stopped at ports of entry. Nevertheless, such persons, for legal purposes, are treated as if they were physically detained outside the country. *See generally* Rosenberg v. Fleuti, 374 U.S. 449 (1963) (discussing when a nonresident alien is considered to have legally entered the United States).

18. The decisions arrived amid fanfare from commentators who viewed the Court's decision as a setback to the Bush Administration. *See* Andrew Buncombe, *Court defies White House with ruling on prisoners*, THE INDEPENDENT (London), June 29, 2004, at 24; Alan Freeman, *Top U.S. Court Undercuts Bush*, THE GLOBE & MAIL (Toronto), June 29, 2004, at A28; Gail Gibson, *Bush Rebuffed on Detention of Terror Suspects*, BALT. SUN, June 29, 2004, at 1A; Charles Lane, *Justices Back Detainee Access to U.S. Courts; President's Powers are Limited*, WASH. POST, June 29, 2004, at A01; Tony Mauro, *Court Affirms Due Process Rights of*

This Article examines two major due process issues left unresolved by the Court: (1) whether nonresident aliens held outside the United States possess due process rights that would entitle them to some form of hearing to challenge their classification as "enemy combatants," and if so, (2) how courts should determine the requirements of those hearings. The analysis, however, is limited to the initial status determination and does not address any subsequent procedures that might be required to justify continued detention of persons properly found to be enemy combatants.[19]

Part I discusses the treatment of various persons, American as well as alien, believed by the executive branch to be associated with Al Qaeda or the Taliban. Part II turns to the question of whether aliens held outside the United States have cognizable liberty interests. Under current procedural due process jurisprudence, the presence of a cognizable liberty or property interest entitles a person to procedural due process.[20] Freedom from confinement is, of course, a liberty interest—perhaps the core liberty interest. But there are lines of cases in the areas of immigration, war powers, and criminal procedure that deny constitutional rights to nonresident aliens outside the country.[21] While many of these cases have been criticized heavily and persuasively by legal scholars,[22] they remain good law in the eyes of the Supreme Court.[23] The goal is to show that those lines of cases do not preclude the recognition of due process rights for persons captured during the war on terrorism because those extraterritorial cases often turn on the difficulty of applying certain Bill of Rights provisions outside the country[24] and involve defendants seeking to be free from our government's action against them rather than seeking the benefit of some action by our government.[25]

Finally, Part III examines the methodology for determining what form of due process a detainee should be given. The Court's recent reflexive citation of the cost-benefit balancing test from *Mathews v. Eldridge*[26] in *Hamdi* is perhaps

---

Enemy Combatants, 231 N.Y. L.J., 2004 (quoting ACLU director Steve Shapiro).

19. For analysis of subsequent procedures, see Tung Yin, *Ending the War on Terrorism One Terrorist at a Time: A Noncriminal Detention Model for Holding and Releasing Guantanamo Bay Detainees*, 29 HARV. J.L. & PUB. POL'Y 149 (2005).

20. *See* Bd. of Regents of State Colls. v. Roth, 408 U.S. 564, 569-70 (1972).

21. *See infra* Part II.B.3-5.

22. *See id.*

23. *See id.*

24. *See infra* Part II.B.5.

25. *See infra* notes 264-81 and accompanying text.

26. 424 U.S. 319 (1976). The test balances "'the private interest that will be affected by the official action' against the Government's asserted interest, 'including the function involved' and the burdens the Government would face in providing greater process." Hamdi v. Rumsfeld, 542 U.S. 507, 529 (2004) (plurality opinion) (quoting *Matthews*, 424 U.S. at 335). The court must then examine "'the risk of an erroneous deprivation' of the private interest if the process were reduced and the 'probable value, if any, of additional or substitute safeguards.'" *Id.* (quoting *Matthews*, 424 U.S. at 335).

understandable,[27] but is also likely to succumb to a result-oriented malleability. The situation in which the interests on both sides can be described with apocalyptic intensity—the individual's freedom versus national security—underscores one of the strongest criticisms of *Mathews*: despite identifying the appropriate factors involved in determining what process is due, the Court has failed to provide any guidance on actually balancing those factors. Therefore, Part III suggests a methodology for determining what process is due. Detainees should be entitled to a hearing to challenge their classification, assistance of and appointment of counsel, and an opportunity to present a defense.

I. BACKGROUND ON THE CURRENT DETAINEE CASES

Identifying the circumstances behind the capture of "enemy combatants" during the war on terrorism is vital to determining whether due process is owed. This Part of the Article examines various persons classified as "enemy combatants" and potentially relevant facts such as citizenship and location of capture. This Part will also examine the actual procedures used to determine their status as enemy combatants.

A. Yaser Esam Hamdi

Yaser Esam Hamdi was captured by Northern Alliance fighters in late 2001 during the U.S.-led military strikes against the Taliban and Al Qaeda in Afghanistan; the Northern Alliance subsequently turned him over to U.S. forces.[28] After initial interrogation, Hamdi was sent to Guantanamo Bay in January 2002, where he stayed until the United States discovered in April 2002 that he held U.S. citizenship and Saudi citizenship.[29] Hamdi was then moved to a Navy brig in Norfolk, Virginia.[30]

Hamdi's father soon learned of his son's detention, and he attempted to secure counsel to represent Hamdi in filing a habeas petition to challenge the detention.[31] The government refused to allow the lawyer, federal public defender Frank Dunham,[32] to meet Hamdi, claiming that Dunham's presence would disrupt the success of ongoing interrogation.[33] The government

---

27. *Id.* at 529.

28. *Id.* at 510.

29. *Id.*

30. *Id.*

31. *Hamdi*, 542 U.S. at 511 (plurality opinion).

32. Katharine Q. Seelye, *Traces of Terror: The Courts; Lawyer Asks for Access to Prisoner Born in U.S.*, N.Y. TIMES, June 21, 2002, at A16.

33. *See* Declaration of Donald D. Woolfolk, *reprinted in* NORMAN ABRAMS, ANTI-TERRORISM AND CRIMINAL ENFORCEMENT app. at 554-55 (1st ed. 2003) [hereinafter Woolfolk Declaration].

subsequently provided Dunham with access to Hamdi but only under restricted conditions and as a matter of grace.[34]

Hamdi eventually challenged his detention under the Fifth and Fourteenth Amendments and the Anti-Detention Act,[35] which prohibits the detention of American citizens "except pursuant to an Act of Congress."[36] The executive branch conceded the Court's jurisdiction to consider Hamdi's habeas petition.[37] But it argued that the President's Commander-in-Chief power was itself a sufficient basis for detaining Hamdi, and suggested that the Congressional authorization for the use of military force against Al Qaeda and the Taliban was an "Act of Congress" that satisfied the Anti-Detention Act.[38] The government also argued that so long as it presented "some evidence" to support its contention that Hamdi was an enemy combatant, the military detention should be justified and, more significantly, no further process should be afforded.[39]

Hamdi's case bounced back and forth between the district court and the Fourth Circuit, with Hamdi consistently prevailing in the former and the government consistently prevailing in the latter.[40] The Supreme Court eventually accepted Hamdi's case for review and ruled in his favor.[41] How big of a victory Hamdi won, however, is less than clear.

In *Hamdi*, the Court failed to reach a majority opinion. Six Justices did vote to remand the case to the lower courts with directions that Hamdi be provided a hearing before a "neutral decisionmaker" in which to challenge his classification as an enemy combatant.[42] But two of the six apparently voted that way merely so that there would be an operative majority decision. Justices Souter and Ginsburg specifically did not join the plurality's discussion of potentially appropriate procedures to be used at an enemy combatant's process hearing, including a possible presumption in favor of the government and use of military judges as the decisionmakers;[43] most likely, Justices Souter and Ginsburg thought that *more* procedures were necessary than those suggested by the plurality.

---

34. Neil A. Lewis, *Sudden Shift on Detainee*, N.Y. TIMES, Dec. 4, 2003, at A1.

35. 18 U.S.C. § 4001 (2000).

36. *Id.* § 4001(a); *Hamdi*, 542 U.S. at 511, 517 (plurality opinion).

37. *See* Brief for the Respondents at 10, *Hamdi*, 542 U.S. 507 (No. 03-6696), 2004 WL 724020.

38. *See Hamdi*, 542 U.S. at 516-17 (plurality opinion). The government also argued that Hamdi did have a process available to contest his classification as an enemy combatant—his interrogation sessions. *See* Transcript of Oral Argument at 30-33, 35-36, 42, *Hamdi*, 542 U.S. 507 (No. 03-6696), 2004 WL 1066082. The Court was not impressed with this argument. *See Hamdi*, 542 U.S. at 526-27 (plurality opinion).

39. *Hamdi*, 542 U.S. at 527-28 (plurality opinion).

40. *See id.* at 511-16 (detailing the lower court holdings).

41. *Id.* at 538-39.

42. *Id.* at 509, 536-39 (plurality opinion); *id.* at 539, 554 (Souter, J., concurring in part, dissenting in part, and concurring in the judgment).

43. *Id.* at 553 (Souter, J., concurring in part, dissenting in part, and concurring in the judgment).

Less than a year after the Court's decision in *Hamdi*, no hearing had taken place, resulting in the government agreeing to release Hamdi from Guantanamo Bay and to deport him to Saudi Arabia, in exchange for, among other things, renouncement of his United States citizenship and his promise not to travel to a list of specified countries (primarily those with an Al Qaeda presence).[44] Therefore, Hamdi's story ends without a clear answer to what form of procedural due process is owed to an American citizen captured overseas and classified as an "enemy combatant."

B. The Guantanamo Detainees

As of September 2006, the United States continues to detain approximately 455 suspected Al Qaeda and Taliban fighters in Guantanamo Bay, Cuba at Camp Delta.[45] While there have been increasing calls for the executive branch to close the Camp Delta detention facility,[46] the government has instead been building new long-term detention facilities.[47] Nevertheless, the due process problems posed by military detention of enemy combatants outside the United States would largely remain despite whether Camp Delta is shutdown because the United States could choose to detain enemy combatants somewhere else outside the country.

What we know of the initial screening process that Camp Delta detainees faced in Afghanistan prior to detainment comes from a speech that Defense Secretary Donald Rumsfeld gave in early 2002.[48] Secretary Rumsfeld explained that those battlefield captures sent to Camp Delta had been designated for transportation to Guantanamo Bay based preliminarily on individual interviews conducted "by a team of people, three or four or five people—sometimes Department of Justice, sometimes Army, mixture of Army, sometimes CIA, sometimes whatever."[49] Secretary Rumsfeld further explained that in some cases, "when doubt [was] raised about [a detainee's status]—[the] process then [was] a more elaborate one" involving additional interviews.[50] These additional interviews or a process of some sort must have taken place at

---

44. *See* Agreement at 2-3, *Hamdi*, No., 2:02-CV-439 (E.D. Va. Sept. 2004), *available at* http://news.findlaw.com/hdocs/docs/hamdi/91704stlagrmnt1.html.

45. *See* DOD News Release, *supra* note 5.

46. *See, e.g.*, Letta Tayler, *Report From Guantamo* [sic]*; Gitmo Truth or Fiction?; Detainees Have Alleged Mistreatment at the Island Facility, but Base Officials Deny the Charges and Say Their Work Makes the U.S. Safer*, NEWSDAY, June 20, 2005, at A20; *Republican Urges Closing Guantanamo Facility*, N.Y. TIMES, June 12, 2005, at 37.

47. *See* Vanessa Blum, *U.S. Building New Prisons for Terrorists: Construction of Guantanamo Jails Signals Long-Term Plans for Base*, LEGAL TIMES, Oct. 4, 2004, at 1; Carol Rosenberg, *Permanent Jail Set for Guantanamo*, MIAMI HERALD, Dec. 9, 2004, at 1A.

48. Donald H. Rumsfeld & General Richard Myers, Pentagon Briefing (Feb. 8, 2002), *available at* http://usinfo.state.gov/regional/nea/sasia/afghan/text/0208dod.htm.

49. *Id.*

50. *Id.*

Guantanamo Bay because the government has released several hundred detainees over the course of the past three years.[51]

In *Rasul v. Bush*,[52] the Supreme Court considered habeas petitions brought by detainees arguing they were not "enemy combatants" and challenging their continued detention at Guantanamo Bay.[53] The Court held that the detainees did have a right under the federal habeas corpus statute to petition for the writ of habeas corpus despite the fact that they were nonresident aliens being detained outside United States territory.[54] That is as far as the opinion goes, however; in effect, the Court held that the detainees had an available legal vehicle through which to vindicate their claims, but the Court said little about the nature of those claims.[55] Subsequently, the Defense Department announced formal hearings aimed at revisiting the combatant status of the detainees.[56] The formal hearings include the assistance of non-lawyer military officers assigned as "personal representatives," the right to call "reasonably available witnesses," and the right to testify (or not to testify).[57]

In the first of the post-*Rasul* detainee cases, *Hamdan v. Rumsfeld*,[58] a Guantanamo detainee slated for prosecution in a military tribunal filed a habeas petition challenging the tribunal's lawfulness. The district court granted the habeas petition but was reversed by the D.C. Circuit.[59] In a 5-3 decision, the Supreme Court invalidated the military tribunals on purely statutory grounds: specifically, the deviance between the procedures used in the military tribunals and those used in courts-martial violated the "uniformity" requirement set forth in the Uniform Code of Military Justice as well as the "regularity" requirement set forth in Common Article 3 of the Geneva Convention.[60] Although *Hamdan* was a notable decision on a number of grounds ranging from its rejection of the President's assertion of unilateral authority to establish military commissions to

---

51. *See* Charlie Savage & Farah Stockman, *Repatriation Of Uzbeks From Guantanamo Halted*, BOSTON GLOBE, June 11, 2005, at A1. The annual "parole-board like" hearings are aimed at determining whether the military should continue detaining a particular detainee classified as an enemy combatant. *Id.*

52. 542 U.S. 466 (2004).

53. *Id.* at 471-72.

54. *See id.* at 478-79.

55. *Id.* at 485 ("What is presently at stake is only whether the federal courts have jurisdiction . . . .").

56. *See* Memorandum from the Deputy Secretary of Defense to the Secretary of the Navy, Order Establishing Combatant Status Review Tribunal (July 7, 2004), *available at* http://www.defenselink.mil/news/Jul2004/d20040707review.pdf.

57. *Id.* at 1-3. For more details on the Combatant Status Review Hearings and *Rasul*, see Tung Yin, *The Role of Article III Courts in the War on Terrorism*, 13 WM. & MARY BILL RTS. J. 1035 (2005).

58. 126 S. Ct. 2749 (2006).

59. 344 F. Supp. 2d 152 (D.D.C. 2004) (mem.), *rev'd*, 415 F.3d 33 (D.C. Cir. 2005), *rev'd*, 126 S. Ct. 2749 (2006). Current Chief Justice Roberts was a member of the D.C. Circuit panel and therefore recused himself from the proceedings before the Court.

60. *Id.* at 2790, 2796.

its holding regarding the applicability of Common Article 3 to suspected Al Qaeda members, it too did not address the question of procedures required—if any—to establish the combatant status of detainees.[61]

Following the district court's ruling in *Hamdan*, two other cases addressed habeas petitions brought by Guantanamo detainees. These cases differed from *Hamdan* in that the habeas petitioners were not challenging the constitutionality of military tribunals, but rather their continued, indefinite detention as enemy combatants.[62] In *Khalid v. Bush*,[63] the district court dismissed the petitions, concluding that while *Rasul* interpreted the habeas statute to provide a vehicle for the detainees to present their claims, it said nothing about the substantive content of those claims.[64] Furthermore, the district court recognized that a long line of Supreme Court precedent denied any constitutional rights to nonresident aliens outside the United States.[65]

Less than two weeks later, the district court in *In re Guantanamo Detainee Cases*[66] reached the opposite result, holding that *Rasul* presupposed the existence of constitutional rights.[67] Because the U.S. naval base was effectively United States territory, the district court recognized a long line of Supreme Court precedent requiring recognition of "fundamental rights" in U.S. territory, including due process.[68] On the merits of the detainees' due process claims, the district court held that the formal hearings failed to provide the detainees with sufficient due process because, among other things, they were denied access to any sort of classified information used to justify their detention and evidence obtained through coercion could be admitted.[69] Both decisions have been appealed and are pending before the D.C. Circuit as of this writing.[70]

Admittedly, the district court's decision in *Khalid* seems less plausibly correct than its decision in *In re Guantanamo*. The writ of habeas corpus is a procedural vehicle for courts to review the legality of the petitioner's custody by the government. To be issued the writ, the petitioner must show that his or her continued custody violates federal law or the Constitution.[71] For the Court to then rule that the Guantanamo detainees have a right to file petitions for writs of habeas corpus while at the same time believing that the detainees have

---

61. *Id.* at 2798 ("It bears emphasizing that Hamdan does not challenge, and we do not today address, the Government's power to detain him for the duration of active hostilities in order to prevent such harm.").

62. *See In re* Guantanamo Detainee Cases, 355 F. Supp. 2d 443, 452-53 (D.D.C. 2005); Khalid v. Bush, 355 F. Supp. 2d 311, 316 (D.D.C. 2005).

63. 355 F. Supp. 2d 311 (D.D.C. 2005).

64. *Id.* at 322.

65. *Id.* at 320-21; *see infra* Part II.B.3-5 (discussing the cases relied upon by the district court).

66. 355 F. Supp. 3d 443 (D.D.C. 2005).

67. *Id.* at 462-63.

68. *Id.*

69. *Id.* at 468, 472.

70. *See* Boumediene v. Bush, No. 05-5062 (D.C. Cir.)

71. Rasul v. Bush, 542 U.S. 466, 473-74 (2004).

no constitutional rights to vindicate through those same habeas petitions seems like a waste of time. In that regard, footnote 15 of Justice Stevens' opinion in *Rasul* offers a tantalizing hint that there may be extraterritorial constitutional rights applicable to the plight of the detainees:

> Petitioners' allegations—that, although they have engaged neither in combat nor in acts of terrorism against the United States, they have been held in Executive detention for more than two years in territory subject to the long-term, exclusive jurisdiction and control of the United States, without access to counsel and without being charged with any wrongdoing—unquestionably describe "custody in violation of the Constitution or laws or treaties of the United States."[72]

Furthermore, Justice Stevens's citation to Justice Kennedy's concurrence in *United States v. Verdugo-Urquidez*,[73] in which Justice Kennedy commented on the realistic extraterritorial application of constitutional provisions,[74] further suggests that Justice Stevens probably envisioned that constitutional rights were available to the detainees.[75]

C. Other Notable Detainees

Many other "enemy combatants," such as Ramzi Binalshibh, Khalid Sheikh Mohammed, Abu Zubaydah, and other unknown "ghost detainees" identified by Human Rights Watch, were detained by the United States in secret locations outside the country,[76] until recently when the government transferred them to Camp Delta.[77] One enemy combatant, Ali Saleh Kahlah al-Marri, is a U.S. resident and Qatari citizen who was initially criminally prosecuted but was subsequently declared an enemy combatant.[78] Unlike the other detainees mentioned here, al-Marri was arrested and held continuously within the United

---

72. *Id.* at 483 n.15 (quoting 28 U.S.C. § 2241(c)(3)).

73. 494 U.S. 259 (1990).

74. *Id.* at 277-78 (Kennedy, J., concurring).

75. *Rasul*, 542 U.S. at 483 n.15. Even so, footnote 15 provides little direction as to which rights are unquestionably violated by the conduct alleged in Rasul's petition. Due process is the obvious choice, but the footnote is not inconsistent with more narrow rights secured by the Geneva Convention if Justice Stevens were to conclude that the Geneva Convention treaty was either self-executing or otherwise provided judicially enforceable rights to individuals. *See generally* Geneva Convention, *supra* note 7 (detailing the rights enjoyed by prisoners of war).

76. HUMAN RIGHTS WATCH, THE UNITED STATES' "DISAPPEARED": THE CIA'S LONG-TERM "GHOST DETAINEES" 8-10, 24-44 (2004) [hereinafter HUMAN RIGHTS WATCH], *available at* http://www.hrw.org/backgrounder/usa/us1004/us1004.pdf.

77. Sheryl Gay Stolberg, *President Moves 14 Held in Secret to Guantanamo*, N.Y. TIMES, Sept. 7, 2006, at A1.

78. Josh Meyer, *Suspect Is Declared an Enemy Combatant*, L.A. TIMES, June 24, 2003, at A1; *see also* Eric Lichtblau, *Bush Declares Student an Enemy Combatant*, N.Y. TIMES, June 24, 2003, at A15.

States.[79] Therefore, the due process analysis relevant to him is different from that of nonresident aliens held outside the country.

Binalshibh, Mohammed, and Abu Zubaydah are all high-level Al Qaeda operatives captured in Pakistan during the war on terrorism.[80] Mohammed allegedly devised the "planes operation," i.e., to have flight-trained hijackers forcibly take control of airliners and crash them into the World Trade Center.[81] Binalshibh, the alleged "20th hijacker,"[82] had visa problems and could not gain entry into the United States to carry out the 9/11 attacks.[83] Instead, he "provide[d] critical assistance from abroad to his co-conspirators" by authorizing money transfers, passing along status reports, and conveying instructions to the 9/11 hijackers from Osama bin Laden.[84]

II. LIBERTY INTERESTS AND DUE PROCESS

Early in the development of due process jurisprudence, the Court drew a distinction between "rights" and "privileges."[85] "Rights," of course, were sacrosanct and thus could not be deprived absent appropriate due process.[86] "Privileges," on the other hand, were merely a matter of government largesse, and because the government had no obligation to provide that largesse, it could revoke privileges without any due process at all.[87] In 1970, the Supreme Court in *Goldberg v. Kelly*[88] did away with the right-privilege distinction.[89] In *Goldberg*, a government agency determined that certain recipients no longer qualified for welfare payments, and the agency decided to terminate the payments without affording the recipients an opportunity for a pretermination hearing to dispute that determination.[90] Under the right-privilege doctrine, the

---

79. *See* Lichtblau, *supra* note 78, at A15.

80. HUMAN RIGHTS WATCH, *supra* note 76, at 26, 31, 37.

81. *See* THE 9/11 COMMISSION REPORT: FINAL REPORT OF THE NATIONAL COMMISSION ON TERRORIST ATTACKS UPON THE UNITED STATES 153-54 (2004) [hereinafter 9/11 COMMISSION REPORT], *available at* http://www.9-11commission.gov/report/911Report.pdf.

82. *See id.* at 168.

83. *Id.*

84. *Id.* at 168, 225, 227, 243-45.

85. *See* BERNARD SCHWARTZ, ADMINISTRATIVE LAW § 5.14 at 254-55 (3d ed. 1991).

86. *See, e.g., id.* at 255.

87. *See, e.g., id.*

88. 397 U.S. 254 (1970).

89. *Id.* at 262 ("The constitutional challenge cannot be answered by an argument that public assistance benefits are 'a "privilege" and not a "right."'" (quoting Shapiro v. Thompson, 394 U.S. 618, 627 n.6 (1969))); *see also* William Van Alstyne, *The Demise of the Right-Privilege Distinction in Constitutional Law*, 81 HARV. L. REV. 1439, 1442 (1968) (noting the "gradual[] ero[sion]" of the right-privilege distinction). *But see* Landon v. Plasencia, 459 U.S. 21, 32 (1982) (holding that aliens seeking admission to the United States were seeking a "privilege").

90. *Goldberg*, 397 U.S. at 256-57.

Court should have dismissed the claim.[91] The Court, however, held that "[r]elevant constitutional restraints apply as much to the withdrawal of public assistance benefits as to" a host of other situations ranging from "denial of . . . tax exemption[s] . . . to discharge from public employment[,]" all of which should trigger due process requirements.[92]

Two years later, however, the Court partially revived the right-privilege distinction in *Board of Regents of State Colleges v. Roth*,[93] holding that when a person has no cognizable liberty or property interests, no procedural rights of any sort are due that person.[94] Roth was hired as an assistant professor under a one-year contract with no guarantee for continued employment.[95] When his contract expired, he was not rehired; the university gave no reason for its decision.[96]

The Court's analysis concerning Roth's lack of a property interest[97] is not particularly relevant to this Article because persons detained in the war on terrorism would be asserting *liberty* interests, not *property* interests. Similarly, the Court's disposal of Roth's asserted liberty interest in being able to teach at his preferred place of employment[98] does not raise concerns, for the Court has always viewed freedom from involuntary confinement as "the core of the liberty protected by the Due Process Clause."[99] Furthermore, the Court has recognized that the default presumption is in favor of such liberty.[100] Therefore, it is not surprising that the Court has repeatedly acknowledged the existence of a detained individual's liberty interest in various cases ranging from civil commitment,[101] to pretrial detention,[102] to deportation proceedings.[103] Nevertheless, *Roth* still suggests that in the absence of a cognizable liberty interest, no process is due to the complaining individual.[104]

---

91. *See, e.g.*, Cafeteria & Rest. Workers Union v. McElroy, 367 U.S. 886, 895, 898 (1961) (recognizing that a "mere privilege" traditionally does not require notice or a hearing and holding that a cook's exclusion from employment for failure to meet established safety requirements did not require notice or a hearing).

92. *Goldberg*, 397 U.S. at 262.

93. 408 U.S. 564 (1972).

94. *Id.* at 569-70.

95. *Id.* at 566.

96. *Id.* at 566-67.

97. *Id.* at 576-78.

98. *Id.* at 575.

99. Foucha v. Louisiana, 504 U.S. 71, 80 (1992) (citing Youngberg v. Romeo, 457 U.S. 307, 316 (1982)), *cited in* Zadvydas v. Davis, 533 U.S. 678, 690 (2001).

100. *See* United States v. Salerno, 481 U.S. 739, 755 (1987) ("In our society liberty is the norm . . . .").

101. *See* Kansas v. Hendricks, 521 U.S. 346, 356-57 (1997) (affirming the involuntary civil commitment of convicted sex offenders).

102. *See Salerno*, 481 U.S. at 755 (affirming pretrial detention)

103. Demore v. Kim, 538 U.S. 510, 523 (2003) (citing *Zadvydas*, 533 U.S. at 697) (affirming detention during deportation proceedings).

104. Bd. of Regents of State Colls. v. Roth, 408 U.S. 564, 569-70 (1972).

Hamdi v. Rumsfeld[105] reaffirmed the principle that freedom from involuntary confinement is a core liberty interest that triggers the Due Process Clause.[106] *Hamdi* also emphasized that the existence of war did not abrogate due process rights.[107] Persons detained in the war on terrorism, apart from whatever physical and mental abuses they might be suffering,[108] are undoubtedly being confined against their will. Therefore, the Court had no problem concluding that Hamdi had a cognizable liberty interest. The Court's recognition of Hamdi's cognizable liberty interest would seem to thereby entitle him to the right to due process, allowing one to move to the question of what process is due.

Yet, the plurality opinion in *Hamdi* referred to the right of *citizens* against arbitrary treatment by their *own* government.[109] This deliberate distinction between citizens and aliens means that the overwhelming majority of persons detained as enemy combatants in the war on terrorism are left without a clear indication as to whether they have due process rights. Furthermore, at the time he filed his habeas petition, Hamdi was physically located within United States territory—a potentially relevant fact, even if it was due to the government's volitional act of transferring him from Guantanamo Bay to Norfolk, Virginia.[110] For those aliens remaining at Guantanamo Bay, detention on what is technically Cuban soil may make their situations distinguishable from Hamdi's.[111]

---

105. 542 U.S. 507 (2004) (plurality opinion).

106. *Id.* at 529.

107. *Id.* at 529-30 ("Nor is the weight on this side of the *Mathews* [*v. Eldridge*, 424 U.S. 319 (1976),] scale offset by the circumstances of war or the accusation of treasonous behavior, for '[i]t is clear that commitment for *any* purpose constitutes a significant deprivation of liberty that requires due process protection . . . .'" (quoting Jones v. United States, 463 U.S. 354, 361 (1983))) (second alteration in original) (emphasis in original).

108. *See* MICHAEL RATNER & ELLEN RAY, GUANTANAMO: WHAT THE WORLD SHOULD KNOW 41 (2004); Matthew Hay Brown, *Freed U.S. Detainees Claim Abuse; The Allegations Could Stir Up Anti-American Sentiment, Already at a Fever Pitch in Islamic Nations*, ORLANDO SENTINEL, Mar. 17, 2004, at A1; Neil A. Lewis, *Frequent Prisoner Coercion Alleged; Guantanamo Personnel Detail Regular Abuse of Those Held at Base*, HOUSTON CHRON., Oct. 17, 2004, at A23; Carol Rosenberg, *3 GIs Punished for Detainee Abuse, Mistreated Inmates at U.S. Military Prison at Guantanamo Bay*, PITTSBURGH POST-GAZETTE, Nov. 12, 2004, at A11; Charlie Savage, *Ex-Detainees Detail Alleged Abuse at US Base*, BOSTON GLOBE, Aug. 5, 2004, at A1.

109. *Hamdi*, 542 U.S. at 531 (plurality opinion) ("We reaffirm today the fundamental nature of a *citizen*'s right to be free from involuntary confinement by his *own* government without due process of law . . . .") (emphasis added).

110. *Id.* at 510-11.

111. This latter point depends heavily on how one reads *Johnson v. Eisentrager*, 350 U.S. 763 (1950) (holding that a nonresident enemy alien has no right of access to U.S. courts in wartime), and *Rasul v. Bush*, 542 U.S. 466 (2004) (holding that federal district courts are not barred from exercising jurisdiction over federal question claims and Alien Tort statute claims asserted by aliens held at Guantanamo Bay). For more, see Yin, *supra* note 57, at 1044-48;

In short, it is possible—even necessary—to narrowly read *Hamdi* as addressing only due process rights of *citizens* being detained on *U.S. territory*, thus leaving unaddressed the due process rights of citizens detained outside of U.S. territory and of aliens altogether. To be fair, *Hamdi* did not present the Court with an opportunity to decide whether aliens—especially nonresident aliens—have cognizable liberty interests and thereby, due process rights. Therefore, Justice O'Connor's focus on citizens can be justified as a matter of judicial prudence. *Rasul v. Bush*,[112] however, did present the opportunity, but the Court did not explicitly address the due process rights of aliens.[113]

Therefore, we must confront whether detainees held outside the United States have cognizable liberty interests and due process rights. Furthermore, because the executive branch has brought United States citizens captured in the war on terrorism within the national borders,[114] this question must focus on aliens whom the United States detains outside the country.

A. The Short-Term Pragmatic Solution

Presently, a pragmatic resolution exists. Despite whether aliens being detained outside the United States have due process rights, one could argue that the U.S. naval base at Guantanamo Bay is, in fact, U.S. territory.[115] Most recently, Judge Green of the D.C. District Court concluded that "[i]n light of . . . *Rasul*, it is clear that Guantanamo Bay must be considered the equivalent of a U.S. territory in which fundamental constitutional rights apply."[116] As Judge Reinhardt noted in *Gherebi v. Bush*,[117] the terms of the lease between Cuba and the United States give Cuba "ultimate sovereignty" over Guantanamo Bay, which he interpreted to mean a "reversionary right" in the event the United States ever vacates the base.[118] In other words, the United States exercises day-to-day sovereignty over the Cuban territory until it leaves. This is evidenced by

---

infra Part II.B.4;

112. 542 U.S. 466 (2004).

113. *See supra* Part I.B.

114. Two U.S. citizens have been classified as enemy combatants: (1) Jose Padilla, who was arrested in Chicago and detained in South Carolina and (2) Yaser Esam Hamdi, who was captured in Afghanistan, brought to Guantanamo Bay and then transferred to Virginia and South Carolina. *See Hamdi*, 542 U.S. at 510-11; Rumsfeld v. Padilla, 542 U.S. 426, 430-32 (2004). Another U.S. citizen, John Walker Lindh, was captured in Afghanistan and brought back to the United States to stand trial on federal charges; he ultimately pleaded guilty pursuant to a plea agreement. United States v. Lindh, 227 F. Supp. 2d 565, 566 (E.D. Va. 2002).

115. *See Rasul*, 542 U.S. at 487 (Kennedy, J., concurring) ("Guantanamo Bay is in every practical respect a United States territory . . . ."); *see also* Gerald L. Neuman, *Closing the Guantanamo Loophole*, 50 LOY. L. REV. 1, 34-44 (2004) (arguing that constitutional rights apply in Guantanamo Bay "under the rationale of the *Insular Cases*").

116. *In re* Guantanamo Detainee Cases, 355 F. Supp. 2d 443, 464 (D.D.C. 2005).

117. 352 F.3d 1278 (9th Cir. 2003).

118. *Id.* at 1293.

the fact that persons, including citizens and aliens, who commit crimes at Guantanamo Bay are brought back to the United States to stand trial.[119]

This argument and calculation, however, is not without complications. The U.S. and Cuban governments presently deny that the land has been annexed by the United States.[120] Therefore, one would think that a debate about the status of the territory would be a paramount example of the sort of political question which courts ought to defer to the political branches of government.[121] Nevertheless, one might avoid the political deferment by arguing that in this particular instance, both the Cuban and United States governments, i.e., the Cuban and United States executive branches, have their own institutional reasons for adhering to the fiction that Cuba retains sovereignty over Guantanamo Bay; Cuba manages to save face in the international community over a matter that continues to frustrate Fidel Castro,[122] while the United States gets to reap the benefits (if any) of extraterritorial detention of enemy combatants.

But this current pragmatic argument and solution might only be of use in the short-term because the executive branch could open a new detention facility in another location that is outside U.S. territory. For example, the United States has previously detained persons captured in Afghanistan or Pakistan in U.S. warships and undisclosed facilities in Afghanistan, Pakistan, Diego Garcia, and Jordan.[123] Unless a new detention facility was located somewhere that, like Guantanamo, can plausibly be said to be the equivalent of U.S. territory, the short-term solution will not succeed and the extraterritorial issue would have to be confronted head on.

Nevertheless, a new detention center would likely be governed by a Status of Force Agreement between the United States and the host nation that might address the rights of persons detained by the United States on such territory. The executive branch, however, would likely forum-shop so as to select the host nation most willing to allow the United States the greatest degree of freedom to operate without oversight.

B. Due Process Outside the United States

Whether a nonresident alien outside the country has due process rights remains an unsettled question, due in part to the existence of conflicting lines of cases. This section discusses the competing doctrines developed in cases dealing with issues ranging from the effect of the Constitution in U.S.

---

119. *Id.* at 1289.

120. *See id.* at 1312 (Graber, J., dissenting).

121. For a more detailed argument on this point, see Yin, *supra* note 57, at 1053.

122. *See Gherebi*, 352 F.3d at 1294 (9th Cir. 2003) (majority opinion) (stating that Cuba has refused to cash the lease checks issued by the United States in protest of the continued presence of American troops on the island).

123. *See* HUMAN RIGHTS FIRST, ENDING SECRET DETENTIONS 9, 15, 16, 17 (2004), *available at* http//: www.humanrightsfirst.org/us_law/PDF/EndingSecretDetentions_web.pdf.

territories and war-time rights. This section then harmonizes the lines of doctrine with an eye toward responding to the parade of horribles identified by Justice Scalia in his dissent in *Rasul*—specifically, that allowing the Guantanamo detainees the right to file habeas petitions (with the implication that those petitions would vindicate applicable constitutional rights) would mean that in World War II, captured German and Japanese prisoners of war would have been entitled to challenge their confinement.[124] Reasonable minds can disagree on the point, but here one can assume that Justice Scalia is correct—it would be unthinkable during a declared war to let thousands, if not potentially millions, of captured prisoners of war, who belong to the armed forces of an enemy nation, bring challenges to their confinement in U.S. courts.

1. The Constitution in Territories

The starting point for discussion of extraterritorial due process is *In re Ross*,[125] an 1891 case in which a U.S. sailor was convicted in a U.S. consular trial for a murder that took place in Japanese waters.[126] Ross had notice of the charges against him, the assistance of counsel, and the right to confront and cross-examine witnesses.[127] He was not indicted, however, by a grand jury and was not tried by a "petit" jury of his peers.[128] Therefore, if he was entitled to the rights available under the Fifth and Sixth Amendments, his trial was unconstitutional.

The Court rejected Ross's claim, holding that "[t]he Constitution can have no operation in another country."[129] Nevertheless, rather than standing for a blanket denial of *any* effect of the Constitution outside U.S. territory, *Ross* can be read more narrowly. In dismissing Ross's claim, the Court noted that, apart from the lack of grand jury indictment and trial by jury, "[i]t is not pretended that the prisoner did not have, in other respects, a fair trial in the consular court."[130] One can read this statement to mean that although not all constitutional rights extended beyond U.S. territory, persons outside U.S. territory are still entitled to fundamental fairness in their trial. After all, the Court would not have upheld a conviction obtained in a court constituted by the United States, even a consular court, where the factfinder flipped a coin to decide the verdict.

Ross was followed by a line of decisions collectively known as the *Insular Cases*,[131] in which the Court confirmed that only "fundamental rights," not

---

124. *See* Rasul v. Bush, 542 U.S. 466, 498 (2004) (Scalia, J., dissenting).

125. 140 U.S. 453 (1891).

126. *Id.* at 454.

127. *Id.* at 458-59.

128. *Id.* at 470.

129. *Id.* at 464.

130. *Id.* at 470.

131. United States v. Verdugo-Urquidez, 494 U.S. 259, 268 (1990) (identifying cases). These five cases are: (1) Balzac v. Porto Rico, 258 U.S. 298 (1922); (2) Ocampo v. United

every constitutional provision, are applicable in unincorporated U.S. territories.[132] For example, the right to trial by jury or indictment by grand jury might not be constitutionally required in territories acquired by the United States.[133] The reasoning for this conclusion was that certain constitutional rights that were not necessarily "fundamental" might be sufficiently alien to the people of unincorporated territories and thus, impractical to impose.[134]

The continued validity of the *Insular Cases*, however, is unclear. In *Reid v. Covert*,[135] the Court declined an opportunity to overrule them.[136] But the Court suggested in dicta that "any further expansion" was unwarranted because "[t]he concept that the Bill of Rights . . . are inoperative when they become inconvenient or when expediency dictates otherwise is a very dangerous doctrine and . . . would . . . undermine the basis of our Government."[137] Furthermore, as Louis Henkin notes, the Court's approach in the *Insular Cases* was similar to that used in the Court's debate about whether the Fourteenth Amendment's Due Process Clause incorporated the Bill of Rights against the States.[138] By the end of the incorporation debate, the Court had concluded that nearly all "of the Bill of Rights [were] . . . 'fundamental' and applicable to the [S]tates" through the Fourteenth Amendment.[139] Henkin speculates whether the Court might also eventually conclude that the same "fundamental" Bill of Rights provisions that apply to the States might also apply in unincorporated territories.[140]

Under the Supreme Court's approach in the *Insular Cases*, due process should be a fundamental right, and therefore available in U.S. territories.[141]

States, 234 U.S. 91 (1914); (3) Dorr v. United States, 195 U.S. 138 (1904); (4) Hawaii v. Mankichi, 190 U.S. 197 (1903); and (5) Downes v. Bidwell, 182 U.S. 244 (1901).

132. *See, e.g.*, *Balzac*, 258 U.S. at 309, 312-13 (stating that "guarantees of certain fundamental personal rights declared in the Constitution . . . ha[ve] . . . full application in the [territories]," but this does not include all of the rights contained in the U.S. Constitution).

133. *See id.*, at 309; *Dorr*, 195 U.S. at 148-49; *Mankichi*, 190 U.S. at 217-18.

134. *See, e.g.*, *Balzac*, 258 U.S. at 310 ("Congress has thought that a people like the Filipinos or the Porto Ricans, trained to a complete judicial system which knows no juries, living in compact and ancient communities, with definitely formed customs and political conceptions, should be permitted themselves to determine how far they wish to adopt this institution of Anglo-Saxon origin, and when.").

135. 354 U.S. 1 (1957) (plurality opinion).

136. *Id.* at 14.

137. *Id.*

138. *See* LOUIS HENKIN, FOREIGN AFFAIRS AND THE CONSTITUTION 308 (2d ed. 1996).

139. *Id.*

140. *Id.*

141. For example, during the Fourteenth Amendment incorporation debates, the Court considered whether particular rights against the federal government contained within the Bill of Rights were "implicit in the concept of ordered liberty." *See* Palko v. Connecticut, 302 U.S. 319, 324-25 (1937). The debate was somewhat abstract since the Due Process Clause of the Fourteenth Amendment served as the focal point for incorporation. But one could safely say that even then, the Court viewed "due process" as implicit in the concept of ordered liberty.

The only issue that remains is whether the Court's reasoning toward unincorporated territories would be applicable for cases arising in foreign countries. In *Reid*, the Court answered this question affirmatively, at least with regard to American citizens.[142] In the earlier case of *Balzac v. Porto Rico*[143] (one of the *Insular Cases*), the Court explained in dicta that "[t]he citizen of the United States living in Porto Rico can not there enjoy a right of trial by jury under the Federal Constitution, any more than the Porto Rican."[144] Thirty years later in *Reid*, however, the Court seemed to reverse itself and held that American citizens overseas (England and Japan) *were* still entitled to constitutional protections.[145] Therefore, the respondents—two women who killed their servicemen husbands—could not lawfully be court-martialed without violating their right to a jury trial.[146] Furthermore, if the Constitution applies overseas for the benefit of American citizens, one might argue that aliens should be similarly protected[147] because the Constitution limits the power of the federal government to act against individuals.[148]

Reid, however, is not sturdy enough to support the weight of that argument itself. The operative passage comes from a plurality opinion, not a majority opinion,[149] and the other Justices who concurred in the result did so largely, if not solely, because the defendants were prosecuted for capital crimes.[150] Furthermore, Justice Black's opinion speaks of limitations on the federal government when acting against *citizens* extraterritorially.[151] His argument that extraterritorial limitations on the power of government are "not a novel concept" draws upon the historical example of Paul, who "successfully invoked

142. *Reid*, 354 U.S. at 5 (plurality opinion).

143. 258 U.S. 298 (1922).

144. *Id.* at 309.

145. *Reid*, 354 U.S. at 5 (plurality opinion).

146. *Id.* In an often-cited passage, Justice Black wrote: "At the beginning we reject the idea that when the United States acts against citizens abroad it can do so free of the Bill of Rights. The United States is entirely a creature of the Constitution." *Id.* at 5-6. This language echoes that from *Balzac*, in which the Court noted that "[i]t is locality that is determinative of the application of the Constitution, in such matters as judicial procedure, and not the status of the people who live in it." *Balzac*, 258 U.S. at 309.

147. *See, e.g.*, Louis Henkin, *The Constitution as Compact and as Conscience: Individual Rights Abroad and at Our Gates*, 27 WM. & MARY L. REV. 11, 24 (1985); Andreas F. Lowenfeld, *U.S. Law Enforcement Abroad: The Constitution and International Law, Continued*, 84 AM. J. INT'L L. 444, 451 (1990); Timothy Lynch, *Power and Liberty in Wartime*, 2004 SUP. CT. REV. 23, 45; Jordan J. Paust, *Antiterrorism Military Commissions: Courting Illegality*, 23 MICH. J. INT'L L. 1, 19 (2001).

148. *Cf.* THE FEDERALIST NO. 84, at 513 (Alexander Hamilton) (Willmoore Kendall & George W. Carey eds., 1966) (arguing that no Bill of Rights was necessary because the Constitution created a federal government of limited and enumerated powers).

149. *Reid*, 354 U.S. at 3, 5 (plurality opinion).

150. *See id.* at 45 (Frankfurter, J., concurring); *id.* at 65 (Harlan, J., concurring).

151. *Id.* at 5 (plurality opinion).

his right as a Roman citizen to be tried in strict accordance with Roman law."[152] Therefore, *Reid* is not inconsistent with the notion that nonresident aliens have constitutional rights. But *Reid* does not compel such a result.

2. Normative Implications of No Extraterritorial Due Process

Before moving past *Reid* and the *Insular Cases*, one should consider the normative implications of holding that nonresident aliens outside the United States do not have due process rights. This conclusion would leave the type of process alien detainees do receive entirely up to the executive branch, without any substantive basis for judicial review. There would be no recourse for relief if the executive branch were to threaten torture or murder detainees. There would even be no recourse for relief if the executive branch were to detain children under the age of 5 as enemy combatants.

Nevertheless, one might hope and expect that the executive branch would face checks and balances in the international political arena; specifically, that international pressure would compel the executive branch to comply with international norms. Peter Spiro has suggested that "[i]nternational actors and international law have . . . been consequential to the resolution of rights-related issues" in the war on terrorism.[153] Writing in 2003, Spiro noted that the Bush Administration had yet to proceed with military tribunals to prosecute suspected alien terrorists despite having authorized them in late 2001, a fact that Spiro attributed to vociferous international opposition.[154]

Spiro took a cautious tone, however, noting "that the episode is still unfolding and that its ending could deviate from the storyline here suggested."[155] His caution was prudent. Despite the opposition from the international community, in 2004 the Bush Administration went ahead with its proposal to conduct war crimes prosecutions of a small number of Guantanamo detainees in the military tribunals authorized under President Bush's November 13, 2001 executive order.[156] Nevertheless, the international political process may yet have a role to play in shaping the executive branch's actions, but that role may only be at the extreme margins, such as deterring the government from harsh and degrading treatment of detainees.[157]

---

152. *Id.* at 6.

153. Peter J. Spiro, *Realizing Constitutional and International Norms in the Wake of September 11*, *in* THE CONSTITUTION IN WARTIME: BEYOND ALARMISM AND COMPLACENCY, at 198, 204 (Mark Tushnet ed., 2005).

154. *Id.* at 204-05.

155. *Id.* at 209.

156. *See* Hamdan v. Rumsfeld, 415 F.3d 33, 35-36, 37 (D.C. Cir. 2005), *rev'd*, 126 S. Ct. 2749 (2006).

157. *Cf.* Gerald L. Neuman, *The Abiding Significance of Law in Foreign Relations*, 2004 SUP. CT. REV. 111, 113 ("The government's resistance to diplomatic overtures underlined the reality that no external authority could constrain its actions at [Guantanamo Bay].").

3. The Plenary Power Doctrine

If the broad language of *Reid* and the related reasoning in the *Insular Cases* provide some possible basis for arguing the extraterritorial reach of the Constitution (even as to aliens), one must confront the line of immigration cases holding that Congress has plenary power to exclude persons from entering the United States and that such excluded persons have no constitutional rights to serve as the basis for challenging their exclusion. The Supreme Court has held that "[a]dmission of aliens . . . is a privilege granted by the sovereign United States Government . . . [which] is granted to an alien only upon such terms as the United States shall prescribe."[158] Furthermore, the Court has recognized that "[w]hatever the procedure authorized by Congress is, it is due process as far as an alien denied entry is concerned."[159] While the right-privilege distinction has fallen by the wayside in all other respects,[160] it retains vitality in the immigration area.[161]

As initially understood, the Plenary Power Doctrine applied to all immigration matters.[162] By the turn of the century, however, the Supreme Court had drawn a significant distinction between deportation, i.e., removal, and exclusion;[163] the alien already in the country who faced deportation also possessed due process rights and was entitled to a hearing to contest deportation.[164] Furthermore, the hearing must comport with the usual requirements of fair notice, opportunity to present evidence, and the right to be represented by counsel.[165] More recently, some adventurous lower courts have made inroads on the Plenary Power Doctrine by requiring a rational basis for exclusion and by examining the conditions of confinement—a due process issue distinct from the decision to exclude itself.[166]

---

158. United States *ex rel.* Knauff v. Shaughnessy, 338 U.S. 537, 542 (1950).

159. *Id.* at 544.

160. *See supra* Part II.

161. *See* Landon v. Plasencia, 459 U.S. 21, 32 (1982).

162. *See* Fong Yue Ting v. United States, 149 U.S. 698, 713-14, 723-24 (1893).

163. Deportation (or removal) proceedings occur when the United States seeks to eject an alien already inside the country; exclusion proceedings, on the other hand, occur when the United States seeks to deny entry to an alien outside the country. THOMAS ALEXANDER ALEINIKOFF ET AL., IMMIGRATION: PROCESS AND POLICY 20, 511 (3d ed. 1995).

164. *See, e.g.*, The Japanese Immigrant Case, 189 U.S. 86, 101 (1903); *see also* Kwong Hai Chew v. Colding, 344 U.S. 590, 596 n.5 (1953) (noting that "once an alien lawfully enters and resides in this country he becomes invested with the rights guaranteed" by the due process clause, which does not recognize "any distinction between citizens and resident aliens"). In addition to aliens inside the country, *resident* aliens (those admitted to lawful permanent residence) may have constitutional rights even when they are outside the country if they are returning to the United States after a short trip abroad. *See Landon*, 459 U.S. at 32-33.

165. *See* The Japanese Immigrant Case, 189 U.S. at 99-101..

166. *See* Stephen H. Legomsky, *Ten More Years Of Plenary Power: Immigration, Congress, and the Courts*, 22 HASTINGS CONST. L.Q. 925, 931-32 (1995). Legomsky notes that "[w]elcome as such decisions are, they are hard to square with the long line of Supreme Court

In the harshest example of the Plenary Power Doctrine, the Court held in *Shaughnessy v. United States ex rel. Mezei*[167] that the indefinite detention on Ellis Island of an alien excluded from entry did not violate the Constitution because the alien was constructively outside the United States, and therefore, had no rights at all.[168] Mezei's situation might be sufficiently analogous to that of the Guantanamo detainees: in both instances, nonresident aliens are detained outside the United States by the United States government for a potentially endless duration. Therefore, like Mezei, the Guantanamo detainees might not have due process rights. The Plenary Power Doctrine and *Mezei* in particular, have been heavily criticized by Supreme Court Justices[169] and legal scholars.[170] Fairly representative of the tenor of criticism is Louis Henkin, who decries the Plenary Power Doctrine as having "emerged in the oppressive shadow of a racist, nativist mood a hundred years ago" and as being "a constitutional fossil, a remnant of a prerights jurisprudence that we have proudly rejected in other respects."[171] Gerald Neuman, another scholarly critic, argues that the Plenary Power Doctrine dates back to old, discredited views about territoriality and sovereignty that faded with *In re Ross*[172].[173]

---

decisions holding due process inapplicable to excluded aliens." *Id.* at 932.

167. 345 U.S. 206 (1953).

168. *Id.* at 207, 213, 215. Although Ellis Island is obviously part of the United States, the fiction at the time was (and remains) that, having been stopped at the border, Mezei was deemed outside the country in the eyes of the law. *Id.* at 215.

169. *See, e.g., id.* at 219 (Jackson, J., dissenting) ("[Mezei], who seems to have led a life of unrelieved insignificance, must have been astonished to find himself suddenly putting the Government of the United States in such fear that it was afraid to tell him why it was afraid of him.").

170. *See, e.g.*, GERALD L. NEUMAN, STRANGERS TO THE CONSTITUTION: IMMIGRANTS, BORDERS, AND FUNDAMENTAL LAW 124-25 (1996); Henry M. Hart, Jr., *The Power of Congress to Limit the Jurisdiction of Federal Courts: An Exercise in Dialectic*, 66 HARV. L. REV. 1362, 1392 (1953) (calling the Plenary Power Doctrine "a patently preposterous proposition"); Stephen H. Legomsky, *Fear and Loathing in Congress and the Courts: Immigration and Judicial Review*, 78 TEX. L. REV. 1615, 1618 (2000) (arguing that the Plenary Power Doctrine is based on an erroneous reliance on federalism cases); Stephen H. Legomsky, *Immigration Law and the Principle of Plenary Congressional Power*, 1984 SUP. CT. REV. 255, 306 (urging the Supreme Court's reexamination of the Plenary Power Doctrine); Michael Scaperlanda, *Polishing the Tarnished Golden Door*, 1993 WIS. L. REV. 965, 1028-31 (arguing that modern developments in international human rights law have seriously "eroded" the Plenary Power Doctrine); Peter H. Schuck, *The Transformation of Immigration Law*, 84 COLUM. L. REV. 1, 18-21 (1984) ("[E]xclusion's extraconstitutional status has encouraged and legitimated some of the most deplorable governmental conduct toward both aliens and American citizens ever recorded in the annals of the Supreme Court.").

171. Louis Henkin, *The Constitution and United States Sovereignty: A Century of* Chinese Exclusion *and Its Progeny*, 100 HARV. L. REV. 853, 862 (1987).

172. 140 U.S. 453 (1891).

173. NEUMAN, *supra* note 170, at 124-25.

Nevertheless, the Plenary Power Doctrine—or at the least, the doctrine adhered to *Knauff v. Shaughnessy*[174] and *Mezei*[175] that nonresident aliens seeking entry in the United States lack constitutional rights—remains resoundingly resilient.[176] Just a few months before the 9/11 attacks, the Court in *Zadvydas v. Davis*[177] relied upon the distinction between deportation and exclusion to note a potential constitutional problem with holding that a federal statute allowed the indefinite detention of aliens who had already been ordered to be removed from the country.[178] The statute in question provided a 90-day period of custody in which to effectuate the removal of the deportable alien; the statute, however, also provided that the alien "may be detained beyond the removal period" in certain circumstances.[179]

The government argued that the statute permitted indefinite detentions of aliens ordered removed if removal was deemed impossible.[180] The Court, however, declined to interpret the statute in such a manner, relying on the maxim that it should avoid interpretation of a federal statute that would "raise[] 'a serious doubt' as to its constitutionality"; therefore, indefinite detention appeared unconstitutional.[181] The Court, discussing the due process available to the petitioners, specifically rejected the government's invocation of *Mezei*, noting that the "critical disctinction" was that Mezei was treated as if he had been stopped at the border while the petitioners had actually reached the country.[182] Therefore, although *Zadvydas* did not actually decide the constitutional questions or repudiate the Plenary Power Doctrine, the Court, in fact, relied on the Doctrine in its analysis of the potential unconstitutionality of the indefinite detention of deportable aliens.[183]

---

174. 338 U.S. 537 (1950).

175. 345 U.S. 206 (1953).

176. *See, e.g.*, David A. Martin, *Graduated Application of Constitutional Protections for Aliens: The Real Meaning of* Zadvydas v. Davis, 2001 SUP. CT. REV. 47, 53.

177. 533 U.S. 678 (2001).

178. *Id.* at 690, 693 ("The distinction between an alien who has effected an entry into the United States and one who has never entered runs throughout immigration law.").

179. 8 U.S.C. § 1231(a)(6) (2000).

180. *See Zadvydas*, 533 U.S. at 685.

181. *Id.* at 689-90 (quoting Crowell v. Benson, 285 U.S. 22, 62 (1932)). Though the case is not clear, the Court appears to be concerned about both the substantive power to detain as well as the procedural aspect of detention. *See id.* at 690-92 (discussing liberty interest in being free from confinement and discussing procedural requirements of noncriminal detention).

182. *Id.* at 693-94.

183. *Clark v. Martinez*, 543 U.S. 371 (2005), also does not signal the demise of the Plenary Power Doctrine. *Clark* involved the same statute at issue in *Zadvydas*, but the statute was being applied to detaining persons ordered excluded rather than deported. *Id.* at 373. The *Clark* opinion was written by Justice Scalia, who had dissented in *Zadvydas*. Scalia wrote that the statute had to be given the same meaning in all circumstances. *Id.* at 378. He had put forward this same rationale in *Zadvydas*. *Zadvydas*, 533 U.S. at 702-05 (Scalia, J., dissenting). Therefore, a statutory interpretation decision, *Clark* is not a decision on what Congress *can* do, but only on what Congress *has* done.

4. The Wartime Cases – *Eisentrager* and *Ludecke*

The strongest precedents against extraterritorial due process rights for nonresident aliens are two wartime cases: *Ludecke v. Watkins*[184] and *Johnson v. Eisentrager*.[185] *Ludecke* involved the constitutionality of the Alien Enemy Act,[186] which authorizes the President, during a congressionally declared war, to round up, detain, and deport those identified as "alien enemies."[187] The Act, enacted in 1798, was invoked by the President on July 14, 1945 to deport over 500 persons.[188] Although Ludecke's status as an alien enemy was determined during proceedings before an Alien Enemy Hearing Board,[189] these proceedings hardly comported with modern notions of due process for several reasons: (1) counsel was usually forbidden[190]; (2) the alien was not permitted to hear the testimony of government witnesses[191]; and most disturbingly, (3) hearing officers deemed "too lenient" by the Justice Department were removed and replaced.[192] Furthermore, "if board members had doubts about an arrested person's loyalty, they generally ruled in favor of the government."[193]

In *Ludecke*, the Court held that the Alien Enemy Act did not violate the Constitution despite its total elimination of judicial review for the Attorney General's decisions and the nature of the hearings involved.[194] Instead, the Court went further, suggesting that the Alien Enemy Hearing Board's hearings were a mere convenience to help the President exercise his statutory authority.[195] One of the dissenting opinions criticized the alien enemy hearings on due process grounds,[196] but the majority opinion failed to even defend the structure of the hearings.[197] The majority instead rested its analysis on "a page of history" relating to the common disposition of alien enemies at the hands of the executive branch "during a state of war."[198]

---

184. 335 U.S. 160 (1948).

185. 339 U.S. 763 (1950).

186. 50 U.S.C. § 21 (2000).

187. *Id.*; *see also Ludecke*, 335 U.S. at 161-62 (citing and discussing the Alien Enemy Act).

188. *Ludecke*, 335 U.S. at 162-63 & n.2. July 1945 is rather late in the game to be deporting German citizens in the United States, but Ludecke had actually been in custody or internment since December 8, 1941. *Id.* at 162-63.

189. *Id.* at 163.

190. *See* ARNOLD KRAMMER, UNDUE PROCESS: THE UNTOLD STORY OF AMERICA'S GERMAN ALIEN INTERNEES 46, 47 (1997).

191. *See id.* at 46.

192. *Id.* at 47.

193. *Id.*

194. *Ludecke*, 335 U.S. at 171.

195. *See id.* at 171-72.

196. *Id.* at 186-87 (Douglas, J., dissenting).

197. *See id.* at 161-73 (majority opinion).

198. *Id.* at 173.

While *Ludecke* dealt with enemy aliens *inside* the United States facing deportation,[199] *Eisentrager* dealt with enemy aliens *outside* the country facing incarceration.[200] The petitioners in *Eisentrager* were Germans who had spied on American forces and passed the information to the Japanese army, despite the fact that Germany had already surrendered.[201] Upon being captured, the petitioners were prosecuted in military tribunals for war crimes and sent back to U.S.-controlled Germany to serve their sentences.[202] The petitioners argued that their military trials violated numerous constitutional provisions,[203] but the Court, per Justice Jackson, rejected this argument.[204] Unfortunately, why the Court did so is unclear.

One can view *Eisentrager* as a strict statutory interpretation decision regarding extraterritorial petitions for writs of habeas corpus. Because the habeas corpus statute limits federal judges to issuing the writ "within their respective jurisdictions,"[205] no court could issue the writ to the petitioners' immediate custodian, who was the commandant of the prison in Germany.[206] The D.C. Circuit Court of Appeals dealt with this problem in *Eisentrager* by holding that the ultimate custodian would be an appropriate respondent if the immediate custodian was outside the territorial jurisdiction of any district court; the Court, however, scoffed at this approach.[207]

Nevertheless, one can also view *Eisentrager* as a constitutional decision recognizing that aliens—or at least enemy aliens outside the United States—lack Fifth or Sixth Amendment trial rights.[208] In explaining why the lower court had erred in granting the habeas writ, Justice Jackson concluded that if certain rights existed extraterritorially, then so did other constitutional rights, which would create a horrific result.[209] "Such extraterritorial application of organic law would have been so significant an innovation in the practice of governments that, if intended or apprehended, it could scarcely have failed to excite contemporary comment."[210]

---

199. *Id.* at 161-63.

200. Johnson v. Eisentrager, 339 U.S. 763, 765-66 (1950).

201. *Id.*

202. *Id.* at 766.

203. *Id.* at 767.

204. *Id.* at 765, 785.

205. 28 U.S.C. § 2241 (2000).

206. *Eisentrager*, 339 U.S. at 766.

207. *See id.* at 781, 785. *Rasul*, it should be noted, effectively nullifies the Supreme Court's interpretation of *Eisentrager* by adopting the lower court's view in that case. Rasul v. Bush, 542 U.S. 466, 483-84 (2004).

208. *See* United States v. Verdugo-Urquidez, 494 U.S. 259, 273-75 (1990) (citing *Eisentrager* and holding that the defendant, a citizen of Mexico facing criminal prosecution in a U.S. federal court, could not invoke the Fourth Amendment to challenge a warrantless search that took place in Mexico).

209. *Eisentrager*, 339 U.S. at 784.

210. *Id.*

Overall, *Eisentrager* can be read as holding that enemy aliens lack both a vehicle for bringing constitutional claims to court and the actual constitutional rights themselves, i.e., the cargo to put in the vehicle. While the former holding is no longer relevant,[211] the latter remains a viable interpretation that stands squarely in the way of an extraterritorial application of the Due Process Clause.

5. The Fourth Amendment Outside the United States

The Court's most recent opinion addressing the extraterritorial reach of the Fourth Amendment was *United States v. Verdugo-Urquidez*.[212] In *Verdugo-Urquidez*, the Court rejected an alien defendant's claim of Fourth Amendment protection with respect to a search of his home in Mexico by U.S. agents even though he was standing trial in the United States and evidence obtained in the search in Mexico was going to be introduced against him.[213] The majority opinion by Chief Justice Rehnquist concluded that "the people" in the Fourth Amendment, as opposed to the use of "person" in the Fifth and Sixth Amendments, suggests that protection against unreasonable searches and seizures should extend "to a class of persons who are part of a national community or who have otherwise developed sufficient connection with this country to be considered part of that community."[214] Therefore, as a nonresident alien brought to the United States involuntarily, Verdugo-Urquidez was not within that class of persons.[215]

The crucial fifth vote for Chief Justice Rehnquist's opinion came from Justice Kennedy, whose separate concurrence essentially tracks the approach of the *Insular Cases*.[216] Justice Kennedy, emphasizing that he was not concluding that nonresident aliens lacked all constitutional rights, stated that nonresident aliens lacked Fourth Amendment protection because applying the Fourth Amendment to extraterritorial searches by U.S. agents would be "impracticable and anomalous."[217] Justice Stevens, who concurred in the judgment, believed that Verdugo-Urquidez did have Fourth Amendment rights but that the search, despite being warrantless, was not unreasonable under the Fourth Amendment.[218] Like Justice Kennedy, Justice Stevens was concerned about the ability of American judges to issue warrants that would be effective in foreign nations and jurisdictions.[219]

---

211. *See Rasul*, 542 U.S. at 483-84.

212. 494 U.S. 259 (1990).

213. *Id.* at 261-62.

214. *Id.* at 265.

215. *See id.* at 274-75.

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

217. *Id.* at 278 (Kennedy, J., concurring). For example, could U.S. judges even issue warrants that would be effective in foreign countries?

218. *Id.* at 279 (Stevens, J., concurring). He was also reassured by the fact that the search was conducted with the cooperation of the Mexican government. *See id.*

219. *See id.*

Justice Stevens and Justice Kennedy were essentially grappling with the problem that some of the Bill of Rights simply do not translate well outside the country. In addition to the search provisions of the Fourth Amendment, foreign application of the Third Amendment prohibition on quartering troops without the consent of the homeowner would apply against U.S. soldiers who take over the home of enemy citizens for bivouacking. Furthermore, local "militias" in occupied territory such as Iraq would have Second Amendment claims not to be stripped of their weapons by the U.S. military if the military wanted to reduce attacks by insurgents.

Foreign application of the Constitution would also cause military prosecutions of enemy soldiers for violations of the laws of war to violate the venue provisions of the Sixth Amendment, which require that the defendant be tried "by an impartial jury of the State and district wherein the crime shall have been committed."[220] If the crime occurred in foreign territory, no *state* could satisfy this requirement.[221] Finally, if the President negotiates a more favorable trade treaty with one country over another similarly situated country due to the perceived intolerance of the second country's dominant religion, the citizens of the second country would arguably have a viable claim of religious discrimination.

6. Resolution

This Article has identified three distinct ways of approaching the extraterritorial constitutional rights of nonresident aliens. First, one can read *Reid v. Covert*[222] and the *Insular Cases*[223] narrowly, the former applying only to U.S. citizens, and the latter applying only within unincorporated U.S. territories.[224] This approach results in the conclusion that nonresident aliens outside the country have no constitutional rights at all. Second, one can argue that the entire Constitution applies everywhere the government acts, a position Gerald Neuman labels "universalism."[225] Third, one could conclude that some, but not all, constitutional provisions apply outside U.S. soil by requiring courts to implement some sort of balancing approach or other methodology.[226] Not one of these three approaches, however, is completely persuasive.

Narrowly reading *Reid* and the *Insular Cases* leads to unquestionably harsh results and seems outdated. The reality of the modern world is that nations exert themselves extraterritorially against aliens directly rather than at the nation-to-nation level.[227] Moreover, the recognition of constitutional limits on

220. U.S. CONST. amend. VI.
221. *See, e.g.*, Johnson v. Eisentrager, 339 U.S. 763, 782-83 (1950).
222. 354 U.S. 1 (1957).
223. *See supra* text accompanying note 131.
224. *See supra* Part II.B.1.
225. NEUMAN, *supra* note 170, at 5-6.
226. *Id.* at 8.
227. Examples abound, but a particularly severe one is the case of Dr. Alvarez-Machain,

government action abroad against citizens, as held in *Reid*, but not against nonresident aliens calls for some explanation of differential treatment. But the most obvious explanations for this differential treatment, such as a social contract theory, would equally justify denial of constitutional protection to aliens (especially illegal ones) inside the United States—a position at odds with current doctrine.[228]

Universalism avoids such potentially xenophobic distinctions but at a different price. Popular with legal scholars but not courts,[229] universalism glosses over some very real problems with applying the Bill of Rights everywhere.[230] Moreover, universalism runs the risk of embroiling federal courts in matters of foreign relations. For example, if nonresident aliens had the full panoply of constitutional rights to assert against the U.S. government, the citizens of a foreign nation could bring a suit alleging that the government's decision to provide aid to another nation but not them violated their Equal Protection and First Amendment rights. A federal court would then invariably be asserting the power to review, and possibly overturn, political decisions by Congress and the President. In *United States v. Curtiss-Wright Export Co.*,[231] the Court rejected this form of judicial review, explaining that "[i]n th[e] vast external realm [of foreign affairs] . . . the President alone has the power to speak or listen as a representative of a nation."[232] Confronted with such a case,

---

who was accused of participating in the torture-murder of U.S. Drug Enforcement Agency agent Enrique Camarena. United States v. Alvarez-Machain, 504 U.S. 655, 657 (1992). Using Mexican locals, the DEA had Alvarez-Machain kidnapped and brought to the United States to stand trial for actions that took place in Mexico but violated U.S. law. *See id.* at 657 & n.2.

228. *See, e.g.*, Plyler v. Doe, 457 U.S. 202, 210 (1982) (holding that illegal aliens were "persons" for the purposes of the Fourteenth Amendment's equal protection guarantee).

229. *See* NEUMAN, *supra* note 170, at 6 ("[T]he universalist approach has significant support among modern commentators, [but] it has played almost no role in American constitutionalism until recent years.").

230. *See supra* Part II.B.4-5.

231. 299 U.S. 304 (1936).

232. *Id.* at 319. Because Congress had already authorized the Presidential action in question, the Court's statement was dicta to the extent that it suggested that the President had an independent foreign affairs power. Nevertheless, *Curtiss-Wright* has assumed an almost mythic status among executive branch lawyers. *See* HAROLD HONGJU KOH, THE NATIONAL SECURITY CONSTITUTION: SHARING POWER AFTER THE IRAN-CONTRA AFFAIR 94 (1990) (reporting that government attorneys refer to the case "as the '"*Curtiss-Wright*, so I'm right" cite'"); H. JEFFERSON POWELL, THE PRESIDENT'S AUTHORITY OVER FOREIGN AFFAIRS: AN ESSAY IN CONSTITUTIONAL INTERPRETATION 23 (2002) (noting that "*Curtiss-Wright* is, understandably, a mainstay in the executive-branch lawyer's kit"). Regardless, the Court has frequently relied on *Curtiss-Wright* for the broader proposition about executive branch power over foreign affairs. *See, e.g.*, Am. Ins. Ass'n v. Garamendi, 539 U.S. 396, 414-16 & n.9 (2003); N.Y. Times Co. v. United States, 403 U.S. 713, 727-28 & n.2 (1971) (*The Pentagon Papers Case*) (Stewart, J., concurring); Hirota v. MacArthur, 338 U.S. 197, 208 (1949) (Douglas, J., concurring); *see also* HENKIN, *supra* note 138, at 20 ("Students of the Constitution may have to accept [Justice] Sutherland's theory, with its difficulties, or leave constitutional deficiencies unrepaired" given

however, the court would almost certainly invoke the political question doctrine to dismiss the case as nonjusticiable.[233] But to say that such matters are nonjusticiable is to say that the plaintiff lacks a *judicially enforceable remedy*, creating a net result that may be the same as holding that the plaintiff lacks a constitutional right: the plaintiff is at the mercy of the executive branch.

To the extent that one endorses the middle approach, which applies some but not all provisions of the Constitution to aliens outside the country, one must confront Justice Black's assertion in *Reid*: "[W]e can find no warrant, in logic or otherwise, for picking and choosing among the remarkable collection of 'Thou shalt nots' which were explicitly fastened on all departments and agencies of the Federal Government by the Constitution and its Amendments."[234] Gerald Neuman's response to Justice Black is the theory of mutuality, whereby aliens are protected by the Constitution "either when they are within the nation's territory or on specific occasions when the nation attempts to exact obedience to its laws."[235] The appeal of Neuman's mutuality approach is that it rests on a reasonable *quid pro quo*: In exchange for being subject to U.S. laws, nonresident aliens outside the country benefit from the constitutional limitations on the government that are part and parcel of the government's power to act in the first place. Neuman's approach, however, stands in direct conflict with *United States v. Verdugo-Urquidez*,[236] and the recognition of possible exigencies that would abrogate some constitutional rights for nonresident aliens may lead to the same functional result as that criticized by Justice Black.[237]

In contrast, Kermit Roosevelt responds to Justice Black's concern by advocating a conflict of laws-based approach to determine which constitutional rights apply extraterritorially by asking "whether its application to a case with foreign elements will promote its domestic purpose."[238] Therefore, Roosevelt suggests that the First Amendment might protect nonresident aliens outside the country from action by our government if the speech in question was with Americans, but not if it was with other aliens.[239] With the Due Process Clause, however, Roosevelt's analysis of government action leads him to ask, "Did we unleash upon the world [a government] with no obligation to respect even the most basic rights of our alien friends?"[240] Hopefully, the answer is no. But

---

that *Curtiss-Wright* "remains authoritative doctrine").

233. *See generally* Baker v. Carr, 369 U.S. 186, 211 & nn.31-33 (1962) (discussing cases that form the precedent for treating issues involving foreign relations as nonjusticiable).

234. Reid v. Covert, 354 U.S. 1, 9 (1957) (plurality opinion).

235. NEUMAN, *supra* note 170, at 108-17.

236. 494 U.S. 259 (1990); *see* NEUMAN, supra note 170, at 108 (noting disagreement with both the Kennedy-Stevens and Rehnquist approaches).

237. *See* NEUMAN, supra note 173, at 115.

238. Kermit Roosevelt III, *Guantanamo and the Conflict of Laws:* Rasul *and Beyond*, 153 U. PA. L. REV. 2017, 2063-65 (2005).

239. *Id.* at 2066-67.

240. *Id.* at 2068.

Roosevelt's question suggests that if the answer is yes, then his conflicts-law based approach would be forced to recognize that nonresident aliens outside the country have no due process rights.[241]

Finally, Diane Amann responds to Justice Black's concern by arguing for the consideration of "external norms," i.e., international human rights laws and customs, as aids in interpreting extraterritorial constitutional rights.[242] Amann argues that international human rights have already grappled with and resolved issues about government responsibility for fundamental human rights, particularly in extraterritorial actions.[243] In fact, Amann's approach is similar to Neuman's,[244] though with greater emphasis on the role of international law. For example, Amann notes that by being captured and controlled by the U.S., the Guantanamo "detainees ha[ve] a structured relationship with the United States, existing over a period of time."[245] Amann's approach, however, requires reliance on international and foreign law in determining the content of U.S. law, an interpretive approach that remains controversial today.[246]

Furthermore, Amann's application of the external norms to the United States' actions in the war on terrorism lead to the conclusion that the United States has violated due process and thus, international law, in a variety of ways including detention without charges, interrogation without counsel, and unfair military commissions.[247] But while international human rights law is well-developed regarding detention without charges, it is much less developed regarding military detention.[248] Amann states that "[d]etention should be based on procedures set out by the legislature, implemented by the executive, and with prompt review by the judiciary."[249] But this analysis overlooks the importance of Congressional military force authorization in the war on

---

241. In that regard, consider one survey soon after the 9/11 attacks: (1) 60% of Americans could envision supporting, as part of the war on terrorism, the assassination of foreign leaders; (2) 32% could support the torture of suspected terrorists; and (3) 27% could see support the use of nuclear weapons. *See, e.g.*, Abraham McLaughlin, *How Far Americans Would Go to Fight Terror*, CHRISTIAN SCI. MONITOR, Nov. 14, 2001, at 1.

242. *See* Diane Marie Amann, *Guantánamo*, 42 COLUM. J. TRANSNAT'L L. 263, 299 (2004).

243. *See id.* at 311-14.

244. *See supra* text accompanying notes 234-37.

245. Amann, *supra* note 242, at 315.

246. *See, e.g.*, Lawrence v. Texas, 539 U.S. 558, 573 (2003) (looking to decisions of the European Court of Human Rights regarding criminalization of homosexual conduct); *id.* at 598 (Scalia, J., dissenting) (arguing that conduct of foreign nations is irrelevant to the content of the Constitution); Atkins v. Virginia, 536 U.S. 304, 316 n.21 (2002) (noting worldwide disapproval of execution of persons with mental retardation); *id.* at 324-25 (Rehnquist, C.J., dissenting) ("I fail to see, however, how the views of other countries regarding the punishment of their citizens provide any support for the Court's ultimate determination.").

247. *See* Amann, *supra* note 242, at 319-35.

248. The Geneva Convention Relative to the Treatment of Prisoners of War, for example, only specifies a status hearing when a detainee's entitlement to prisoner of war status is in doubt.

249. Amann, *supra* note 242, at 322.

terrorism, which presumably includes the substantive power to detain enemy combatants.[250]

Amann's approach would require the United States to provide due process—judicial hearings, counsel, and so forth—to prisoners of war captured in a traditional nation-state war. Even the most analogous international law, the Geneva Convention Relative to the Treatment of Prisoners of War, does not call for any hearings to justify detention except when the detaining authority seeks to strip prisoners of POW status if the prisoner's lack of entitlement to such status is in doubt.[251] The current war on terrorism does not involve a nation-state enemy, a difference that may be significant enough to make the traditional war model inapplicable. But this merely means that the external norms of international law may not have anything sufficiently concrete to say about due process when the targets of domestically authorized military force are non-state actors who are not insurgents and spontaneous defenders identified in the Geneva Convention.

To raise concerns about these various approaches, however, is not to deny their ingeniousness. But from a practical, doctrinal perspective, there is something to be said for solving the particular problem at hand—in this instance, the due process rights, or lack thereof, of nonresident aliens detained in connection with Congressionally authorized military force.[252] Unlike the First, Second, Third, Fourth, Fifth, or Sixth Amendments, the Due Process Clause does not present conceptual difficulties regarding extraterritorial application. The question of scope is, of course, different, and the question may lead to the conclusion that the liberty interests of aliens involuntarily confined by our government overseas result in lesser process than if they were confined within the country or were citizens.

The Due Process Clause speaks of "persons,"[253] and from a textual perspective, the use of "person" arguably leads to broad coverage unless one demonstrates that nonresident aliens outside the country are not "persons." A contrary conclusion would lead to some truly curious results. For example, a nonresident alien would not be a "person" outside the United States, but would

---

250. Indeed, her article does not even cite, much less discuss, the significance of the Authorization for Use of Military Force. *See* Authorization for the Use of Military Force, *supra* note 2.

251. Geneva Convention, *supra* note 7, art. 5.

252. *See* DANIEL A. FARBER & SUZANNA SHERRY, DESPERATELY SEEKING CERTAINTY: THE MISGUIDED QUEST FOR CONSTITUTIONAL FOUNDATIONS 152 (2002) (noting that "the common law method" of "develop[ing] and elaborat[ing] principles" as one goes along, instead of "attempting to articulate a general theory from the start," can lead to "satisfyingly reasoned resolutions in difficult cases"). Of course, Farber and Sherry were critiquing the effort by such scholars as Robert Bork, Richard Epstein, and Bruce Ackerman to develop their own grand unifying theory covering all aspects of constitutional law. *See id.* at 152-53. Neuman and Roosevelt are at the most, articulating grand theories covering a particular aspect of constitutional law. *See supra* text accompanying notes 233-41.

253. U.S. CONST. amend. XIV, § 1.

magically become a "person" as soon as he or she set foot inside the United States.[254] Furthermore, when a nonresident alien acquired resident status, he or she would be "elevated" to the status of a "person"—a rather nation-centric approach.

The only entities that the Court has explicitly held not to be "persons" for Due Process Clause purposes are fetuses[255] and States.[256] The methodology the Court used in those cases involved an examination of the Constitutional provisions mentioning "person," which resulted in the conclusion that "person" could not be used to encompass those entities.[257] That same methodology, however, does not compel the conclusion that nonresident aliens are not "persons" under the Due Process Clause. The broadest reading of the Due Process Clause would simply extend all Fifth Amendment individual protections—indictment, double jeopardy, self-incrimination— to nonresident aliens in appropriate circumstances. Even then, a court could plausibly draw distinctions between the purposes of the various Fifth Amendment protections so that due process applies despite whether all provisions do. For example, corporations are "persons" for due process purposes,[258] but do not have the right against self-incrimination.[259]

Therefore, as Ronald Dworkin asserted, "if noncitizens across the world have any due process protection against our government at all," it is the right "not to be arbitrarily and indefinitely imprisoned."[260] By focusing on the sole issue of indefinite detention, Dworkin avoids having to confront many of the contrary lines of authority discussed above.[261] Again, the *Insular Cases*[262] could only be read to apply the Constitution to U.S. territories, but it is surely not inconsistent with them to recognize due process as a "fundamental right" that the government is obligated to respect everywhere. Furthermore, Dworkin's narrow focus on indefinite detention neatly elides the immigration cases, which do not involve indefinite detention.[263]

Overall, it helps to distinguish between different types of liberty interests, i.e., one can conceptualize liberty in two ways: (1) being left alone by the government and (2) asking for the government's help. The Court has made this distinction in the substantive due process context. In *DeShaney v. Winnebago County Department of Social Services*,[264] the Court rejected DeShaney's claim

254. *See supra* Part II.B.3.

255. *See* Roe v. Wade, 410 U.S. 113, 158 (1973).

256. *See* South Carolina v. Katzenbach, 383 U.S. 301, 323-24 (1966).

257. *See, e.g., id.*

258. *See* Santa Clara County v. S. Pac. R.R. Co., 118 U.S. 394, 394 (1886).

259. Braswell v. United States, 487 U.S. 99, 102 (1988).

260. Ronald Dworkin, *What the Court Really Said*, N.Y. REV. OF BOOKS, Aug. 12, 2004, *available at* http://www.nybooks.com/articles/17293.

261. *See supra* Part II.B.1-5.

262. *See supra* text accompanying note 131.

263. *See supra* Part II.B.3.

264. 489 U.S. 189 (1989).

that social workers and other government officials "deprived him of his liberty" by failing to remove him from his father's custody despite knowledge of complaints of parental abuse, which resulted in massive brain damage.[265] In concluding that DeShaney failed to allege a due process violation, the Court drew on reasoning from *Harris v. McRae*,[266] stating: "'Although the liberty protected by the Due Process Clause affords protection against unwarranted *government* interference . . . , it does not confer an entitlement to such [governmental aid] as may be necessary to realize all the advantages of that freedom.'"[267] Because DeShaney asked the government to protect him from someone else's action rather than affirmative government action against him, the Court concluded that he had no constitutional liberty interest at stake.[268]

Detainees involuntarily confined by the government in a prison camp thousands of miles away from their homes and families obviously present the clearest example of persons asking to be left alone by the affirmative action of the government. Their freedom to move about, to work, and to spend time with family members and friends are all curtailed by confinement. If they were American citizens, they would have due process rights despite whether they were detained at Guantanamo or elsewhere outside the United States.[269] If they were aliens detained *inside* the United States, they would have due process rights.[270] Nonresident aliens are simply asking for an end to the government's *affirmative action* against them.

In the alien exclusion cases, however, the Court has justified the denial of any constitutional rights to nonresident aliens seeking entry to the country on the ground that permission to enter is a *privilege*,[271] i.e., the nonresident alien, by seeking admission to the United States, wants to be helped by the government and does not want to be left alone. The aliens might believe, for example, that their lives would be better in the United States than their home country because of a greater freedom to enjoy the fruits of liberty. But the United States, under the Court's reasoning, is under no obligation to assist the alien in realizing that greater freedom.[272] In other words, those aliens are asking for an end to the government's *inaction* rather than action towards them.

The various possibilities of extraterritorial application of liberty interests no doubt fall on a spectrum rather than a bright-line division. One difficult case,

---

265. *Id.* at 191.

266. 448 U.S. 297 (1980).

267. *DeShaney*, 489 U.S. at 196 (quoting *Harris*, 448 U.S. at 317-18) (emphasis in original) (alteration in original).

268. For criticism of *DeShaney*, see Laurence H. Tribe, *The Curvature of Constitutional Space: What Lawyers Can Learn From Modern Physics*, 103 HARV. L. REV. 1, 8-14 (1989).

269. *See* Reid v. Covert, 354 U.S. 1, 12, 14 (1957) (plurality opinion).

270. *See, e.g.*, Zadvydas v. Davis, 533 U.S. 678, 688-90 (2001).

271. *See, e.g.*, Landon v. Plasencia, 459 U.S. 21, 32 (1982); *see also* HENKIN, *supra* note 138, at 307 n.**.

272. *Cf. supra* notes 266-67 and accompanying text.

for example, is *Shaughnessy v. United States ex rel. Mezei*.[273] In *Mezei*, an alien was "excluded from the United States on [national] security grounds."[274] When efforts to send him to Great Britain, France, Hungary, and "about a dozen Latin-American countries" failed, he informed U.S. officials that "he would exert no further efforts to depart."[275] The Court held that his continued, indefinite detention at Ellis Island did not violate the Due Process Clause.[276] Mezei, however, was essentially asking to be left alone by our government,[277] and yet, the Court rejected the notion that he had any cognizable liberty interest. Therefore, *Mezei* stands as a counter-example to the taxonomy identified above.

Another way to think about *Mezei* is to recognize that the United States would have been delighted to end Mezei's detention had he been willing and able to find another country that would have accepted him. This fact is in direct contrast to the situation of the Guantanamo detainees, who remain confined even though many of their home countries have pressed for their release. Therefore, Mezei's case presents an almost unique situation, as he was asking for help, i.e., entry into the United States, and was denied, but because no other country would accept him and releasing him would admit him into the country, the only option was to indefinitely detain him on Ellis Island.[278]

Admittedly, this "unique situation" rationale smacks of the discredited right-privilege distinction, but it is also akin to the action-inaction distinction that is well-recognized in criminal and tort law.[279] Although it does not resolve all questions involved in extraterritorial application of the Due Process Clause,[280] it is at least consistent with the prevailing view that if the Due

---

273. 345 U.S. 206 (1953).

274. *Id.* at 207. Mezei was actually a resident alien at one time, but left the United States and sought reentry, a fact that resulted in his being treated functionally "as if stopped at the border." *Id.* at 213, 215.

275. *Id.* at 209.

276. *Id.* at 215.

277. *Cf. id.* at 220 (Jackson, J., dissenting) ("Realistically, this man is incarcerated by a combination of forces which keep him as effectually as a prison, the dominant and proximate of these forces being the United States immigration authority.").

278. This conclusion, however, should recognize the harshness of the result in Mezei's case, and one can plausibly argue that Mezei was not just asking for help, but instead through circumstances beyond his control, he was unable to get away.

279. *See generally* WAYNE R. LAFAVE, CRIMINAL LAW § 3.3(a) (3d ed. 2000) (discussing the general rule that "one has no legal duty to aid another person in peril, even when that aid can be rendered without danger or inconvenience to himself").

280. For example, consider the controversial practice of rendition, whereby the United States deports an individual to a country known to use torture as an interrogation method. As agents of that country torture the individual for information, U.S. agents listen passively. Does this taxonomy help us decide whether the person is entitled to due process under our Constitution? Perhaps not on the theory that the American agents are only there to listen and lack the power to stop the interrogation. But perhaps yes, on the theory that the American agents really are in control of the process.

Process Clause has a primary purpose, that purpose is to prevent the government from arbitrarily detaining an individual involuntarily and indefinitely.[281]

C. The Enemy Alien Disability Rule

The view that citizens *and* aliens have cognizable liberty interests regardless of where they are being held by our government might seem incompatible with the executive branch's power to wage war. This incompatibility becomes apparent if one asks, as Justice Scalia did,[282] whether enemy soldiers captured during a traditional nation-state war (World War II) have a cognizable liberty interest not to be unjustly detained as prisoners of war.[283]

One potential rejoinder is that the mere existence of a cognizable liberty interest does not automatically mean that the person will be entitled to the procedures he or she asks for; the liberty interest merely means that the person is entitled to have the court determine what constitutionally required procedures would minimize the risk of an erroneous deprivation of that interest. On the other hand, to allow captured enemy soldiers to litigate the scope of procedures within American courts in an attempt to demonstrate that they are being detained wrongfully seems incongruous. Even if the courts ultimately conclude that no procedures beyond those implemented unilaterally by the executive branch are required, the very subjugation of the executive branch's military decisions to the courts could "hamper the war effort and bring aid and comfort to the enemy."[284]

---

281. *See, e.g.*, Neuman, *supra* note 115, at 44 ("[I]f the Due Process Clause addresses any extraterritorial conduct, it surely speaks to the calculated establishment of an offshore prison, outside any war zone or territory under belligerent occupation, where captives can be gathered from afar for prolonged detention, interrogation, and possible execution."); Dworkin, *supra* note 260.

282. *See* Rasul v. Bush, 542 U.S. 466, 488-99 (2004) (Scalia, J., dissenting).

283. Posing the question this way may seem more provocative than necessary. One can conclude, of course, that enemy soldiers have a cognizable liberty interest, but that the interest is not outweighed by the government's interest in waging war effectively. Alternatively, one could conclude that such a liberty interest exists, but that no additional procedures are warranted to protect the interest.

284. Johnson v. Eisentrager, 339 U.S. 763, 779 (1950). Gerald Neuman has criticized this aspect of the *Eisentrager* opinion, noting that modern habeas practice allows a habeas petition to be resolved as a matter of law without the need to produce the petitioner before the court. Neuman, *supra* note 115, at 61. Furthermore, Neuman notes that teleconferencing and other technological solutions can be used to address Justice Jackson's concerns. *Id.*

1. The Enemy Alien Disability Rule in Traditional Wars

Both *Johnson v. Eisentrager*[285] and *Ludecke v. Watkins*[286] can be understood to establish the rule that during times of war, enemy aliens, both enemy soldiers and enemy civilians, are denied access to United States courts if granting access would interfere with the executive branch's ability to effectively wage war.[287] Whether these enemy aliens have an applicable vehicle for getting to court, such as the federal *habeas corpus* statute, or whether they have constitutional claims to press in court, such as a due process right does not matter. If armed conflict exists between the United States and the enemy alien's home nation, the enemy alien is *disabled* from accessing our courts because that access could hamper the war effort. As Justice Jackson explained in *Eisentrager*, "[D]isabilities this country lays upon the alien who becomes also an enemy are imposed temporarily *as an incident of war and not as an incident of alienage*."[288] Therefore, the denial of court access ends when the armed conflict does.[289]

The enemy alien disability rule has a long history, and in the past, it has been applied quite harshly. For example, alien citizens of nations at war with each other have not only been denied access to courts, but their property was also subject to forfeiture.[290] In more recent times, the rule has been relaxed. Enemy aliens do not automatically lose claims to assets, and if an enemy alien won a lawsuit prior to the outbreak of war but judgment had not been paid, the defendant could still be ordered to pay the judgment to an Alien Property Custodian, who would hold the judgment in escrow until after the end of the war.[291]

---

285. 339 U.S. 763 (1950).

286. 335 U.S. 160 (1948).

287. *See Eisentrager*, 339 U.S. at 776, 778 ("But the nonresident enemy alien . . . does not have . . . access to our courts . . . ."); *Ludecke*, 335 U.S. at 171-73.

288. *Eisentrager*, 339 U.S. at 772 (emphasis added).

289. Given that *Eisentrager* was decided in 1950, five years after Germany surrendered, one might wonder why the petitioners in that case were still subject to the enemy alien disability rule. The technical, and perhaps unsatisfactory, answer is that the state of war between the United States and Germany did not officially end until October 19, 1951, by a joint resolution of Congress. Joint Resolution to Terminate the State of War Between the United States and the Government of Germany, Pub. Law No. 181, 65 Stat. 451 (1951).

290. *See, e.g.*, Clarke v. Morey, 10 Johns. 69, 72, 74 (N.Y. Sup. Ct. 1813) (outlining the history of the practice in the United States and Europe).

291. *See* Birge-Forbes Co. v. Heye, 251 U.S. 317, 323 (1920); *see also* Hanger v. Abbott, 73 U.S. 532, 536 (1867) (holding that debt held by an alien would be suspended until the war but not forfeited); Russel v. Skipwith, 1 Serg. & Rawle 310, 310 (Pa. 1815) (holding that if an enemy alien had a valid cause of action prior to the outbreak of hostilities, "the plea of alien enemy, which is never favored, merely suspends the action, and on the removal of the disability, it revives in full force").

Furthermore, the scope of disability does not apply to all cases brought by enemy aliens. In *Ex parte Kawato*,[292] a Japanese citizen residing lawfully in the United States brought a suit during World War II to recover wages and other compensation he claimed he was entitled to; the employer argued that the suit was barred under the enemy alien disability rule.[293] After tracing the gradual relaxation of the enemy alien disability rule from its roots in English common law to the post-Revolutionary American version, the Court concluded that the modern form of the disability rule applied against *resident* alien enemies "only so far as necessary to prevent use of the courts to accomplish a purpose which might hamper our own war efforts or give aid to the enemy."[294] Kawato's lawsuit, the Court reasoned, would not impact the U.S. war effort, but instead, simply required another private party to pay him wages and other monies.[295]

Kawato is not an unreasonable narrowing of the rule. If enemy aliens still residing inside the United States at the time of military conflict are capable of hindering the war effort through litigation or other method, they can be deported pursuant to the Alien Enemy Act.[296] If they have not been deported, the executive branch obviously perceives these aliens as being "enemy aliens" as a consequence of birth, and not as a descriptive matter of conduct.[297]

Nevertheless, *Kawato* might be read as narrowing the enemy alien disability rule in civil cases between enemy aliens and other *non-government* parties. In bolstering its conclusion that Kawato should have been allowed to proceed with his suit, the Court noted that "[i]n asking that the rights of resident aliens be abrogated in their behalf, private litigants in effect seek to stand in the position of government."[298] If this is the correct interpretation of *Kawato*, then enemy aliens remain barred from bringing even civil lawsuits against the United States government. For example, in *Harisiades v. Shaughnessy*,[299] the Court reaffirmed that the United States had no obligation to pay just compensation for property taken from enemy aliens during wartime.[300] Despite this interpretation of *Kawato*, the Court held in *Ex parte Quirin* that German saboteurs captured in New York in 1942 during World War II were entitled to seek Supreme Court review of their military tribunal convictions,

---

292. 317 U.S. 69 (1942).

293. *Id.* at 70-71.

294. *Id.* at 72-75.

295. *Id.* at 78; *see also* Clark v. Allen, 331 U.S. 503, 505-06, n.1, 508, 510 (1947) (striking down a California statute prohibiting enemy aliens from inheriting property because a treaty with Germany granted such a right and the treaty had not been repealed nor had Congress passed contrary legislation).

296. 50 U.S.C. § 21 (2000).

297. *See Kawato*, 317 U.S. at 71 ("'Alien enemy' as applied to petitioner is at present but the legal definition of his status because he was born in Japan, with which we are at war.").

298. *Id.* at 74.

299. 342 U.S. 580 (1952).

300. *Id.* at 586 n.9 (citing Russian Volunteer Fleet v. United States, 282 U.S. 481 (1931)).

over the government's argument that they should have been denied access to our courts.[301] But it is one thing to deny enemy soldiers affirmative access to our courts; it is another to deny them a right to make a defense at a military tribunal, including legal challenges to the jurisdiction of a tribunal.[302]

This narrow reading of *Quirin*, however, raises a conflict with *Johnson v. Eisentrager*[303] because one could also characterize the petitioners in that case as challenging the jurisdiction of the tribunal over them. The Court itself perceived no conflict in *Eisentrager*. The Court concluded that in *Quirin* and the case before it, it had merely provided the enemy aliens with opportunities to show why the enemy alien disability rule should not have been applied against them, and that in both cases, the enemy aliens failed to persuade the Court.[304] If the same conclusion was reached in both cases, i.e., "that no right to the writ of *habeas corpus* appears,"[305] then *Quirin* and *Eisentrager* are not in conflict.

But *Eisentrager* offers a revisionist reading of *Quirin*. In *Quirin*, the Court stated the issue under review as whether "the Constitution and laws of the United States constitutionally enacted forbid [an enemy alien's] trial by military commission."[306] The Court did ultimately deny the German saboteurs leave to file petitions,[307] lending some support to *Eisentrager*'s reading of the case; but *Quirin* also held "that the [military] Commission ha[d] jurisdiction to try the charge preferred against [the] petitioners,"[308] suggesting a decision on the merits. Therefore, either (1) *Quirin* represents an odd exception to the enemy alien disability rule, whereby enemy aliens *inside* the United States can defend themselves against military prosecutions but enemy aliens *outside* the United States cannot,[309] or (2) *Quirin* did not reject the enemy alien disability rule but still resolved the merits of the claim against the petitioners.[310]

The enemy alien disability rule is not without controversy. Part of the rule is dependent on geography, which harkens back to the old notion of territoriality—a restriction that Kermit Roosevelt has described as

---

301. *See Ex parte* Quirin, 317 U.S. 1, 24 (1942).

302. *See id.* at 25 ("But there is certainly nothing in the Proclamation to preclude access to the courts for determining its applicability to the particular case."); *see also In re* Yamashita, 327 U.S. 1, 9 (1946) (citing *Ex parte* Quirin, 317 U.S. 1, 24, 25 (1942) for the same proposition).

303. 339 U.S. 763 (1950).

304. *Id.* at 780-81.

305. *Id.* at 781 (emphasis in original).

306. *Ex parte* Quirin, 317 U.S. at 25.

307. *Id.* at 48.

308. *Id.* at 25.

309. One might rationalize this distinction on the ground that if the executive branch voluntarily brings enemy soldiers into the country, it cannot complain if they are granted access to our courts. But that view is held by a nation confident that it will not be fighting on its own territory.

310. *See Ex parte* Quirin, 317 U.S. at 25 ("[W]e have resolved [the merits] . . . against petitioners. There is therefore no occasion to decide contentions of the parties unrelated to this issue.").

"demonstrably false," full of "a lengthy list of defects," and "not . . . an appealing model."[311] In particular, Roosevelt wonders "why it should make a difference that the government has (fortuitously or strategically) elected to hold [enemy aliens] outside its borders."[312] One answer might be that during a traditional war such as World War II, we might expect to capture thousands, if not millions, of enemy soldiers. Therefore, there should be a default position that blocks enemy soldiers from our courts and that allows the executive branch to wage war without judicial supervision. Roosevelt's insight, however, is not without force. It might make more sense to deny enemy soldiers access to our courts regardless of where they are held during the duration of the military conflict.[313] Allowing court access to enemy soldiers held within the country may be an artifact of *Quirin* and the old notions of territoriality, but the existence of such an artifact does not mean that the entire rule should be discarded.[314]

2. Inapplicability to the War on Terrorism

The enemy alien disability rule is easy to apply in the traditional armed conflicts between nation-states, particularly the armed forces of the enemy nation.[315] The laws of war require soldiers to wear distinctive uniforms precisely so that combatants can be easily identified and distinguished from civilians.[316] As a result, enemy soldiers in compliance with the laws of war will have no basis to challenge the application of the enemy alien disability rule.

Furthermore, there is no reason to think that the enemy alien disability rule will lead to unduly harsh results in the context of nation-state wars. One can imagine that someone detained as an enemy soldier might claim to be a noncombatant. The enemy soldier might even claim that he found the uniform

---

311. Roosevelt, *supra* note 238, at 2044, 2046.

312. *Id.* at 2046.

313. *See id.* at 2045 ("It is not unimaginable that the purposes of the Constitution would be better served by denying some protections to . . . those who have entered the country . . . to make war on it or those we have deemed enemies and brought within our borders for incapacitation or punishment.") (citation omitted).

314. *But see* Amann, *supra* note 242, at 290-92 (arguing that the enemy alien disability rule is not only "anachronistic" but has been undermined by the due process revolution of the mid-20th Century).

315. The rule is even easy to apply to enemy civilians. Under the Alien Enemy Act, 50 U.S.C. § 21 (2000), for example, aliens will unlikely dispute their citizenship with the enemy nation; rather, they are more likely to dispute their imputed loyalty to the enemy nation. *See* KRAMMER, *supra* note 190, at 48 (noting an appeal from a German enemy alien proclaiming loyalty to the United States during World War II). But disputing loyalty may be irrelevant, because under the Alien Enemy Act, all that matters is the existence of war between the United States and the alien's nation, as well as the alien's citizenship with that enemy nation. 50 U.S.C. § 21 (2000). Everything else is a matter of the President's discretionary implementation of the statute. *Id.*

316. *See* Geneva Convention, *supra* note 7, art. 4.A.(2)(b).

on a dead soldier and put it on because he was cold, and that he picked up the weapon to protect himself from wild animals. But the practical reality is that a hearing, even if available, would be of no use to such a person. What witnesses would verify the detainee's claims? The obvious answer is someone in the enemy nation's military command, but during a traditional war, the United States is hardly in a position to subpoena a witness from the enemy.

A person might also claim that he is an American soldier who escaped from enemy capture and donned the enemy uniform to avoid detection by enemy soldiers.[317] Again, requiring a hearing to address this unlikely event seems like overkill. The detainee would just give his American name, rank, and serial number and identify his commanding officer or fellow soldiers; these factual assertions could be checked without the need for any formal process.

Even absent a uniform, there are situations where a detainee's combatant status should be apparent by his or her own actions. Article 4 of the Geneva Convention recognizes that some combatants will not be wearing uniforms because such a requirement would exclude "organized resistance movements" and locals "who on the approach of the enemy spontaneously take up arms to resist the invading forces."[318] Article 4, however, also requires that nonprofessional fighters identify themselves as combatants by at least carrying their arms openly, and, in the case of organized resistance groups, "having a fixed distinctive sign recognizable at a distance."[319] This requirement reduces the likelihood of civilian casualties by making combatants clearly identifiable—one of the goals of the modern laws of war.[320] The 1979 Protocol Additional to the 1949 Geneva Conventions provide some functional revisions to this requirement, but still retain the ultimate goal of making combatants easily identifiable.[321] Therefore, combatants retain lawful combatant status if they carry weapons openly "[d]uring each military engagement and . . . [d]uring such time as he is visible to the adversary while he is engaged in a military deployment preceding the launching of an attack in which he is to participate."[322]

The enemy alien disability rule also does not alter substantive rights, including, in particular, those provided by the Geneva Convention. Therefore, enemy prisoners of war are denied access to our courts to challenge, for example, the conditions of their confinement, and, in the event of war crimes

317. While putting on an enemy uniform might be considered "perfidy," the American soldier will not have violated the laws of war as long as he does not attack the enemy while in the stolen uniform. *See* Protocol Additional to the Geneva Conventions of 12 Aug. 1949, and relating to the Protection of Victims of International Armed Conflicts, art. 37, Dec. 7, 1979, 1125 U.N.T.S. 3 [hereinafter Protocol 1].

318. Geneva Convention, *supra* note 7, art. 4.A.(2), (6).

319. *Id.* at art. 4.A.(2).

320. *See, e.g.*, George H. Aldrich, *New Life for the Laws of War*, 75 AM. J. INT'L L. 764, 765 (1981).

321. Protocol 1, *supra* note 317, art. 44.

322. *Id.* art. 44.3.(a).

prosecutions, the manner of such prosecutions. The United States, however, is obligated by the Geneva Convention to meet minimum standards for detention and military commissions.[323] Although no court has ultimately held that the Geneva Convention is self-executing,[324] the treaty does contemplate that violations by a signatory will be redressed at the international level.[325]

The current war on terrorism, however, presents a different, murkier situation. There is no enemy nation, and although a state of armed conflict does exist between the United States and Al Qaeda,[326] membership in Al Qaeda is not likely to be so apparent. Some captured members, such as Zacarias Moussaoui and Richard Reid, may freely admit their ties to Al Qaeda, but others will deny such ties.[327] An additional complication is that persons who have been or are still being detained at Guantanamo Bay as suspected Al Qaeda fighters include citizens from longstanding allies such as Great Britain and Australia.[328]

The significant due process issue involved in the war on terrorism is whether the detainee is in fact a member of Al Qaeda or the Taliban, a classification that would justify detention under the laws of war. Therefore, application of the enemy alien disability rule would cause unduly harsh results. Moreover, the argument that witnesses needed to corroborate a detainee's claim to being a noncombatant are unavailable due to the state of war is not persuasive in the context of the war on terror. For example, there may be Al Qaeda members still at large in Afghanistan or Pakistan, but that does not mean that non-Al Qaeda or Taliban persons in those countries are enemies of the U.S.—presently those nations are friendly to the U.S. Furthermore, the consequence of the enemy alien disability rule during a traditional nation-state war is that the enemy alien detainee remains confined for the duration of the war. With the war on terrorism, however, the "duration of the war" takes on a

---

323. *See, e.g.*, Geneva Convention, *supra* note 7, arts. 17, 25 (prohibiting threats, torture, or any "unpleasant or disadvantageous treatment of any kind" as interrogation tactics); and requiring that "[p]risoners of war shall be quartered under conditions as favourable as those for the forces of the Detaining Power who are billeted in the same area").

324. *See, e.g.*, Hamdan v. Rumsfeld, 415 F.3d 33, 39-40 (D.C. Cir. 2005).

325. *See* Geneva Convention, *supra* note 7, art. 132.

326. *See* Bradley & Goldsmith, *supra* note 2, at 2109.

327. The government has frequently alleged that Al Qaeda members are trained to resist interrogation and, more specifically, to abuse the American legal system. Desmond Butler, *German Judge Frees Qaeda Suspect; Cites U.S. Secrecy*, N.Y. TIMES, Dec. 12, 2003, at A3; Neil A. Lewis, *Ashcroft Defends Antiterror Plan; Says Criticism May Aid U.S. Foes*, N.Y. TIMES, Dec. 7, 2001, at A1; *see also* A. John Radsan, *The Moussaoui Case: The Mess from Minnesota*, 31 WM. MITCHELL L. REV. 1417, 1436 (2005) (arguing that Moussaoui should be treated as an enemy combatant because his self-representation during his criminal trial provides a blueprint to other Al Qaeda fighters on how to abuse the U.S. legal system).

328. *See, e.g.*, Raymond Bonner, *Australia Uneasy About U.S. Detainee Case*, N.Y. TIMES, Apr. 10, 2005, § 1, at 10. One of the detainees, David Hicks, slated for trial before a military commission is Australian; and numerous detainees, including the lead petitioner in *Rasul v. Bush*, are British.

whole new meaning; from the standpoint of the detainee, detention could be a life sentence. Therefore, during a traditional war, the enemy alien disability rule distributes its burden on the individual and his home nation, which can obtain the release of its captured citizens by agreeing to end the armed conflict. But the same rule during a war against non-state actors concentrates the burden solely on the individual, who may be unable to end the armed conflict.[329]

III. DETERMINING WHAT PROCESS IS DUE

Once a claimant has shown the existence of a cognizable liberty interest, one must determine what process is due that person to ensure that the liberty or property interest is not erroneously infringed. As noted earlier, *Rasul v. Bush*[330] did not reach this issue,[331] so there has been no official pronouncement with regard to nonresident aliens *outside* the country. Nor did *Hamdan v. Rumsfeld*[332] decide the matter, as that case invalidated the President's military commissions on purely statutory grounds. *Hamdi v. Rumsfeld*,[333] however, provides some guidance because in that case, the Court addressed the due process of citizen-detainees.[334] Having concluded that Hamdi was entitled to *some* due process, the Court turned to *Mathews v. Eldridge*[335] for "[t]he ordinary mechanism that we use for balancing such serious competing interests, and for determining the procedures that are necessary to ensure that a citizen is not 'deprived of life, liberty, or property, without due process of law.'"[336]

But why should *Mathews* be the clear choice for due process methodology? Drawing upon familiar critiques of the *Mathews* balancing test, this Part shows that *Mathews* is a particularly ill-suited approach for determining appropriate due process when the interests at stake on the part of the individual and the

---

329. *But see* Yin, *supra* note 19, at 196-201 (suggesting additional process to determine if detainees remain dangerous to justify continued detention).

330. 542 U.S. 466 (2004).

331. *See supra* Part I.B.

332. 126 S. Ct. 2749 (2006).

333. 542 U.S. 507 (2004) (plurality opinion).

334. *Id.* at 537. The Court also addressed the "threshold question" of whether the President could lawfully detain an American citizen as an enemy combatant, but the Court disposed of that issue by reference to the congressional authorization to use military force and *Ex parte Quirin*, 317 U.S. 1 (1942). *Id.* at 516-18. Justices Souter and Ginsburg refused to join this portion of Justice O'Connor's opinion. *Id.* at 539 (Souter, J., concurring in part, dissenting in part, and concurring in the judgment), but the plurality could count on Justice Thomas, who argued in dissent that the President had the authority to detain Hamdi even in the absence of the congressional authorization, based on the power as Commander-in-Chief. *Id.* at 579-81 (Thomas, J., dissenting).

335. 424 U.S. 319 (1976).

336. *Hamdi*, 542 U.S. at 528-29 (plurality opinion) (quoting U.S. CONST. amend. V). In determining the process due to the nonresident alien Guantanamo detainees, subsequent to *Hamdi*, a district court cited *Hamdi* and also applied the *Mathews* balancing test. *See In re* Guantanamo Detainee Cases, 355 F. Supp. 2d 443, 465-67 (D.D.C. 2005).

government can be expressed in extreme terms: freedom from confinement versus national security. This Part then argues that a more appropriate methodology is to determine the core components of due process required in analogous detention situations (prison disciplinary hearings, civil commitment proceedings, and the like). If a core component exists across the spectrum, there arises a presumption that the same component exists with respect to detention of persons in the war on terrorism.

A. The Trouble with Mathews

One might have expected some explanation of the appropriateness of using the *Mathews* balancing test in the context of the war on terrorism. It might seem odd to use a test developed in an administrative law case dealing with a property interest in discretionary government transfer payments.[337] The Court merely cited five earlier cases without further explanation of their applicability to this situation. It is true that these cases involved liberty interests based on involuntary detention. In two of the cases, the Court used the *Mathews* balancing test to uphold the constitutionality of procedures governing pretrial detention.[338] In the other cases, the Court used *Mathews* to evaluate procedures governing civil commitment of the mentally ill.[339] Perhaps the Court could have reasoned that *Mathews* offered the presumptively correct approach to analyzing due process claims, and alternative approaches were no more suitable for determining appropriate due process in the war on terrorism.

In *Mathews*, a recipient of discretionary Social Security disability payments was notified that the state agency charged with monitoring his condition had determined that he was ineligible for continued benefits, and the Social Security Administration accepted this conclusion.[340] The claimant argued that, under *Goldberg v. Kelly*,[341] he was entitled to an "evidentiary hearing" with nearly full trial-like procedures prior to terminating his disability benefits.[342] Along the way to distinguishing *Goldberg*, the Court set forth its now familiar three factor test:

---

337. *See* Lassiter v. Dep't of Soc. Servs., 452 U.S. 18, 59-60 (1981) (Stevens, J., dissenting) (arguing that *Mathews* should be limited to cases involving property interests; but in a case involving both liberty and property interests, "[t]he issue is one of fundamental fairness, not of weighing the pecuniary costs against the societal benefits"); Edward L. Rubin, *Due Process and the Administrative State*, 72 CAL. L. REV. 1044, 1137 (1984) (arguing prior to use of *Mathews* in nonadministrative cases that such use "would produce major dislocations in existing doctrine, since it would be very difficult to derive the components of criminal or civil trials from that framework").

338. United States v. Salerno, 481 U.S. 739, 746 (1987); Schall v. Martin, 467 U.S. 253, 274-75 (1984).

339. Heller v. Doe, 509 U.S. 312, 330-31 (1993); Zinermon v. Burch, 494 U.S. 113, 127-28 (1990); Addington v. Texas, 441 U.S. 418, 425 (1979).

340. Mathews v. Eldridge, 424 U.S. 319, 324 (1976).

341. 397 U.S. 254 (1970).

342. *Mathews*, 424 U.S. at 325.

> [D]ue process requires consideration of three distinct factors: First, the private interest that will be affected by the official action; second, the risk of an erroneous deprivation of such interest through the procedures used, and the probable value, if any, of additional or substitute procedural safeguards; and finally, the Government's interest, including the function involved and the fiscal or administrative burdens that the additional or substitute procedural requirement would entail.[343]

According to the Court, the private interest here—the claimant's need for an uninterrupted stream of disability benefits—was potentially "significant" but "likely to be less than that of a welfare recipient."[344] Thus, relative to *Goldberg*, the claimant in *Mathews* had a lesser claim for the same kinds of procedures. Next, in considering the marginal value of the requested procedure (in this case, a pretermination evidentiary hearing), the Court concluded that disability benefits cases were different from welfare benefits cases in that the former relied primarily on medical reports, rather than assessments of the personal veracity of recipients regarding efforts to find work.[345] For that reason, oral testimony, cross-examination, and other such rights were likely to be of little to no value in the disability cases.[346] Finally, the Court concluded that the government's interest in avoiding the not insubstantial cost of providing pretermination evidentiary hearings, combined with an interest "in conserving scarce fiscal and administrative resources," was sufficient to outweigh the claimant's desire for such hearings.[347]

Common criticism of *Mathews* concerns the Court's failure to articulate exactly how the three factors are to be balanced against one another.[348]

---

343. *Id.* at 335.

344. *Id.* at 342.

345. *Id.* at 343-44.

346. *Id.* at 344-45.

347. *Id.* at 348.

348. JERRY L. MASHAW, DUE PROCESS IN THE ADMINISTRATIVE STATE 106 (1985) (describing the language in the *Mathews* opinion as "loose"); Rubin, *supra* note 337, at 1138 (noting that "it is unclear how to resolve the inevitable conflicts between" the three factors, and that the cost-benefit approach is "not a promising one"). Broader criticism of *Mathews* goes to its single-minded focus on cost-benefit efficiency, thereby ignoring other values that, while not related to accuracy, nevertheless play important roles in validating government action. Jerry Mashaw notes, for example, that the right to vote is worthless "in practical, political terms," but that disenfranchisement would nevertheless strike us as unfair, because "[i]nvolvement in the process of political decision making . . . seems to be valued for its own sake." MASHAW, *supra*, at 163; Richard B. Saphire, *Specifying Due Process Values: Toward a More Responsive Approach to Procedural Protection*, 127 U. PA. L. REV. 111, 155 (1978) ("[T]he utilitarian balancing of the first two elements against a broadly-defined governmental interest would inevitably dilute or eliminate dignitary protection."). Mashaw argued that a model of due process should incorporate a "dignitary theory," such that participants feel they have been treated fairly, without regard to the accuracy-improvement of the additional process. MASHAW, *supra*, at 162-64, 177-80; *see also* Frank I. Michelman, *Formal and Associational Aims in Procedural Due Process*, *in* DUE PROCESS: NOMOS XVIII 127-28 (J. Roland Pennock & John

Mathews states merely "that identification of the specific dictates of due process generally requires consideration of three distinct factors."[349] No guidance is given as to how one weighs these three factors. If the government's interest is especially weighty, does an individual have to show that the requested procedure would greatly increase the accuracy of the results? Conversely, if the individual's interest is especially weighty, will a marginal increase in accuracy suffice?

Given this uncertainty, some scholars have conceptualized *Mathews* as calling for a cost-benefit analysis of the form:

(increase in accuracy) x (claimant's interest) > (burden on government)[350]

A utilitarian cannot help but be drawn to the factors identified by *Mathews*. If due process is flexible, thus implying *some* form of balancing, these appear to be the correct factors to consider.[351] The individual may have reasons to prefer to use certain procedures, and the government may have reasons to prefer not to use those same procedures, and why bother with procedures that do not improve the accuracy of the factfinding?

Obviously, this type of formula works best with property interests that can be translated into money, and even more so if the government's interest can also be translated into financial terms. Unfortunately, federal courts are unlikely to come up with the quantitative data that the cost-benefit approach demands.[352] In *Mathews* itself, the Court conceded an inability to calculate the precise cost to the government of the additional or substitute procedure.[353]

W. Chapman, eds., 1977).

349. *Mathews*, 424 U.S. at 335.

350. STEPHEN G. BREYER ET AL., ADMINISTRATIVE LAW AND REGULATORY POLICY 843 (5th ed. 2002); *see also* Parrett v. City of Connersville, 737 F.2d 690, 696 (7th Cir. 1984) (noting that *Mathews* "merely requires a comparison of the costs and benefits of giving the plaintiff a more elaborate procedure than he actually received"); Sutton v. City of Milwaukee, 672 F.2d 644, 645 (7th Cir. 1982) (describing the *Mathews* test as "a simple cost-benefit test"); Propert v. District of Columbia, 741 F. Supp. 961, 962 (D.D.C. 1990) (characterizing *Mathews* as calling for "a cost-benefit" analysis); Lawrence B. Solum, *Procedural Justice*, 78 S. CAL. L. REV. 181, 252-53 (2004) (describing the *Mathews* test as a "consequentialist version if imperfect procedural justice").

351. *See* RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 563-64 (6th ed. 2003) (Judge Posner notes that the *Mathews* formula looks very similar to the Hand Formula used to conceptualize the reasonableness of precautions taken against reasonably foreseeable harm under tort law); *see also* United States v. Carroll Towing Co., 159 F.2d 169, 173 (2d Cir. 1947) (describing the Hand Formula).

352. *See* MASHAW, *supra* note 348, at 126 (explaining that the Court "needs information on the causes and incidence of errors, [and] the social valuation of both positive and negative errors, the error-correcting power of various procedural devices"); Saphire, *supra* note 348, at 155 (questioning the "assumption that the societal costs of procedural protection can reliably be predicted").

353. *Mathews*, 424 U.S. at 347.

With respect to Social Security disability pretermination hearings, however, the Court was satisfied that the increase in costs would be substantial, if for no other reason than that it was guaranteed that virtually every claimant would invoke the "attractive option" of a right to such a hearing.[354] Furthermore, the Court concluded that the threat to the individual's interest was unlikely to be as severe as that in welfare cases, supposedly because persons on welfare were already on subsistence monies, whereas persons on Social Security disability may have had other forms of income available.[355]

In subsequent cases, the Court has used a similar "guesstimate" approach when implementing the *Mathews* balancing test.[356] A representative case is *Walters v. National Ass'n of Radiation Survivors*, where the Court upheld a $10 limitation on fees that could be paid out of veterans' benefits to attorneys.[357] In rejecting the argument that the fee limitation infringed the due process right to be represented by counsel (because few, if any, lawyers would take such cases for fees capped at $10), the Court applied *Mathews* and concluded that the presence of counsel was highly unlikely to increase the accuracy of the veterans' benefits payouts.[358] The Court conceded that it was "[im]possible to quantify the 'erroneous deprivations' among the general class of rejected claimants,"[359] and while the Court did make some attempt to review statistics regarding the success rate of appeals based on method of representation,[360] this attempt merely sampled the success rates without making efforts to determine if the different samples were truly comparable.

The problem with applying the *Mathews* balancing test becomes even more acute when one examines liberty interests. How does one weigh an individual's interest in not being confined indefinitely on Guantanamo Bay? How much weightier does that interest become if one considers the fact that detention on Guantanamo Bay is accompanied by coercive interrogation and other forms of mental and physical abuse? How much less weighty does that interest become if the government shows that the detainee has been eating more and better food than he had been before his capture?[361]

Of course, the criticism that the *Mathews* factors cannot be quantified with mathematical precision can be overstated. With regard to the individual's

---

354. *Id.*

355. *Id.* at 342.

356. *See, e.g.*, Walters v. Nat'l Ass'n of Radiation Survivors, 473 U.S. 305 (1985).

357. *Id.* at 307-08, 335.

358. *Id.* at 334.

359. *Id.* at 327.

360. *Id.*

361. One might argue that this analysis takes too broad a view of the government's interests beyond the fiscal cost to the government of the additional procedures requested. *See* Mathews v. Eldrige, 424 U.S. 319, 347 (1976) (noting that the public interest involves "the administrative burden and other societal costs" associated with the additional procedure). If this is the proper reading of *Mathews*, however, it leads to the rather odd result that the balancing test fails to take into account nonfiscal interests, such as national security.

interest at stake, one might not be able to quantify the exact value that the person places on the particular liberty or property interest, but one may well be able to ordinally rank that interest against other liberty or property interests. Therefore, one can say with a high degree of confidence that a person's liberty interest in being free from confinement is greater than the person's property interest in, for example, practicing a line of work.[362] Accordingly, one might say that if certain procedures have been found necessary under the given property interest, then those procedures are the minimum for the given liberty interest (assuming, of course, that the government's interest remains the same).

But ordinal ranking is not likely to be of much use in determining due process in the war on terrorism. The individual's interest is concededly quite high, even more so than that in *Goldberg*, where the claimant was on the edge of subsistence.[363] But the government's interest will undoubtedly be viewed as quite high, as well. After all, the government has "national security" at stake, which the Court has accepted as a "compelling" interest when used to justify civil liberties infringements,[364] and has, in fact, upheld government actions that it might not otherwise have upheld but for the claim of national security.[365]

---

362. In determining the significance of the interest at issue in *Walters* (the radiation survivor case), the Court analogized the benefits to those in *Mathews*, rather than the subsistence payments in *Goldberg*. *Walters*, 473 U.S. at 333.

363. Indeed, it would only be higher if the death penalty were at issue.

364. *See, e.g.*, Selective Serv. Sys. v. Minn. Pub. Interest Research Group, 468 U.S. 841, 860-61 (1984) (Powell, J., concurring); *see also* Jean v. Nelson, 472 U.S. 846, 872 (1985) (Marshall, J., dissenting) (characterizing earlier cases as holding that due process was not violated despite denial of hearing for persons seeking entry to the country, based on government's national security interest).

365. *See* Nat'l Treasury Employees Union v. Von Raab, 489 U.S. 656, 674 (1989) (upholding drug testing of Secret Service employees against privacy claims despite fact that only 5 out of 3600 tested positive "[i]n light of the extraordinary safety and national security hazards that would attend the promotion of drug users to positions that require the carrying of firearms or the interdiction of controlled substances"); Snepp v. United States, 444 U.S. 507, 509 n.3 (1980) ("Moreover, this Court's cases make clear that—even in the absence of an express agreement—the CIA could have acted to protect substantial government interests by imposing reasonable restrictions on employee activities that in other contexts might be protected by the First Amendment."); Branzburg v. Hayes, 408 U.S. 665, 728 n.4 (1972) (Stewart, J., dissenting) ("In Zemel v. Rusk, 381 U.S. 1 [(1965)], we held that the Secretary of State's denial of a passport for travel to Cuba did not violate a citizen's First Amendment rights. The rule was justified by the 'weightiest considerations of national security' and we concluded that the 'right to speak and publish does not carry with it the *unrestrained* right to gather information.'"); Korematsu v. United States, 323 U.S. 214, 219 (1944) (upholding exclusion of tens of thousands of Japanese-Americans from the West Coast due to a claim of national security during World War II); Near v. Minnesota, 283 U.S. 697, 716 (1931) (suggesting in dicta that the First Amendment would allow a prior restraint to be issued against "publication of the sailing dates of transports or the number and location of troops" and other such critical wartime information); United States v. Progressive, Inc., 467 F. Supp. 990, 996 (W.D. Wis. 1979) (rejecting a First Amendment defense against prior restraint to prevent publication of a magazine article providing technical guidance for building a nuclear weapon).

It will help to think about the two directions in which error can lie. Recall that the individual's interest at stake is a reduction in the risk of erroneous deprivation of liberty or property—that is, a false positive. On the other hand, the government's interest at stake, in addition to fiscal and administrative concerns,[366] is a reduction in the risk of an erroneous windfall of liberty—that is, a false negative.[367] Thus, the individual detainee wants procedures that will maximize his opportunity of showing that he is not a member of Al Qaeda or the Taliban. The government, on the other hand, wants procedures that will minimize the likelihood that an actual member of Al Qaeda or the Taliban will be mistakenly released. The fact that there are two dimensions of error is not always apparent from the Court's decisions; for example, in *Lassiter v. Dep't of Social Services*, the Court referred to the government's "interest in a correct decision."[368] Yet, in *Mathews*, the Court was at least cognizant of the possibility that someone might receive disability benefits undeservedly and that the government had an interest in recovering such payments.[369]

In any event, the lack of guidance as to how to weigh the *Mathews* factors could lead a court to take a careful look at the improved accuracy of requested procedures and to conclude that they: (1) would lead to a greater increase in false negatives than decrease in false positives and (2) given the government's paramount interest in protecting the public,[370] the procedures are unwarranted. As discussed in more detail below, the presence of an attorney could, under such an analysis, be deemed as not improving the overall accuracy of the results and be excluded as a constitutionally required procedure, despite its use in analogous circumstances. Indeed, perhaps the greatest indictment against the use of the *Mathews* balancing test is that in *Hamdi v. Rumsfeld*, in concluding that Yaser Hamdi was "unquestionably" entitled to counsel, Justice O'Connor did not resort to the *Mathews* test but simply asserted it as a given.[371] This stands in stark contrast to her tentative thoughts that the hearing to which Hamdi was entitled might permit the use of hearsay, as well as a presumption in

366. *See, e.g.*, City of Los Angeles v. David, 538 U.S. 715, 718 (2003) (per curiam) (rejecting the requirement of immediate hearings on towed cars because "administrative resources available to modern police departments are not limitless"); Washington v. Harper, 494 U.S. 210, 232 (1990) (holding that requirement of *judicial* hearings would divert critical resources) (emphasis added).

367. In a related vein, consider Justice Harlan's concurrence in *Winship*, in which he provided a sophisticated analysis of the "beyond a reasonable doubt" standard of proof as a mechanism for allocating "the relative frequency of these two types of erroneous outcomes." *In re* Winship, 397 U.S. 358, 370-71 (1970) (Harlan, J., dissenting).

368. 452 U.S. 18, 31 (1981).

369. *See* Mathews v. Eldridge, 424 U.S. 319, 347 (1976) (discussing the "theoretical right of the Secretary to recover undeserved benefits"). *But see* MASHAW, *supra* note 348, at 123-24 (arguing that "the Court does not identify the costs resulting from the positive errors induced by pretermination hearings").

370. *See* Mackey v. Montrym, 443 U.S. 1, 17 (1979) ("We have traditionally accorded the states great leeway in adopting summary procedures to protect public health and safety.").

371. 542 U.S. 507, 539 (2004) (plurality opinion).

favor of the government, both of which were justified within the context of *Mathews* balancing.[372] If such an obvious entitlement as access to counsel (at least, for a citizen-detainee) need not be justified by reference to *Mathews*, perhaps the test itself is the problem.

B. It's Not All About Mathews

To the extent that *Mathews* retains a seductive appeal, one should keep in mind that it does not provide the only way of determining the process that is due in a given situation. The Court has held in a variety of contexts that some approach other than the *Mathews* balancing test is appropriate for determining whether a given procedure satisfies due process.[373] Thus, whether notice provided is proper is determined under the "reasonableness" test of *Mullane v. Central Hanover Trust Co.*,[374] and military procedures are evaluated under *Middendorf v. Henry*'s deferential standard without regard to cost-benefit balancing.[375]

Most significantly, in *Medina v. California*,[376] the Court held that "the *Mathews* balancing test does not provide the appropriate framework for assessing the validity of state procedural rules which . . . are part of the criminal process."[377] In the Court's view, the various constitutional rights set forth in the Bill of Rights (particularly the Fourth, Fifth, and Sixth Amendments) represent a "careful balance . . . between liberty and order" that could be disturbed by the use of *Mathews* to apply the "open-ended rubric of the Due Process Clause."[378] While this reasoning might have led the Court to conclude that only criminal procedures mandated by the Bill of Rights need be considered, the Court did not go so far. Instead, the Court concluded that the appropriate test was to consider whether the failure to adopt the requested procedure would "offend[] some principle of justice so rooted in the traditions and conscience of our people as to be ranked as fundamental."[379]

Indeed, a great deal of criminal procedure could not plausibly fit within any conception of the *Mathews* balancing test.[380] The right against self-

372. *Id.* at 533-34.

373. *See, e.g.*, Middendorf v. Henry, 425 U.S. 25 (1976); Mullane v. Central Hanover Trust Co., 339 U.S. 306 (1950).

374. *See* Dusenbery v. United States, 534 U.S. 161, 167-68 (2002) (citing *Mullane*, 339 U.S. at 313).

375. *See* Weiss v. United States, 510 U.S. 163, 177-78 (1994) (stating that *Middendorf* calls for the adoption of the procedures requested by the claimant only where the factors in favor of the procedures "are so extraordinarily weighty . . . to overcome" Congress's determination of the default procedures in the Code of Uniform Military Justice (quoting *Middendorf*, 425 U.S. at 43-44)).

376. 505 U.S. 437 (1992).

377. *Id.* at 443.

378. *Id.*

379. *Id.* at 446 (quoting *Patterson v. New York*, 432 U.S. 197, 202 (1977)).

380. This is not to say that criminal procedure involves no balancing and only bright line

incrimination, the exclusionary rule, and attorney-client (and other) privileges all generally *restrict* the factfinders from considering relevant, probative evidence,[381] and therefore *decrease* the accuracy of the proceedings.[382] Nevertheless, these procedures are all well-accepted parts of the criminal justice system.

This is not to suggest that constitutional criminal procedure is the appropriate methodology for determining the process that is due to the Guantanamo detainees, for the detainees do not face criminal prosecution. Because due process is flexible, what is appropriate for the liberty deprivation setting of criminal prosecution may not be appropriate for another liberty deprivation setting. Rather, it is to suggest that just as pure cost-benefit analysis and accuracy-determination do not drive the criminal procedure, they might not be the sole factors to consider in determining the content of due process in noncriminal detention settings.

C. The Core Due Process Requirements in Detention Cases

If the *Mathews* balancing test is not the right approach for determining what level of process is due to detainees held in the war on terrorism, what is the correct approach? If due process is flexible, as the Court has stated on numerous occasions,[383] then there must be some balancing involved. As Richard Fallon succinctly explains, "No process is ever perfect, and in determining how many procedural guarantees to require, the Supreme Court needs to weigh interests in fairness to individuals against the government's interest in being able to make decisions swiftly, without excessive costs of time or money."[384]

Generally speaking, due process balancing results in the use of one of three different approaches, as summarized by William Buss: (1) "specify a set of procedures and make them applicable to all, or at least a broad range, of cases"; (2) "work out the due process calculus for each case"; or (3) "defer to the

---

rules. *See, e.g.*, New York v. Quarles, 467 U.S. 649, 656-57 (1984) (balancing suspect's *Miranda* rights against public safety); United States v. Martinez-Fuerte, 428 U.S. 543, 555 (1976) (discussing weighing of public interest against Fourth Amendment interests).

381. *See, e.g.*, Thomas C. Grey, *Procedural Fairness and Substantive Rights*, *in* DUE PROCESS: NOMOS XVIII, *supra* note 348, at 183.

382. *See, e.g.*, Massiah v. United States, 377 U.S. 201, 208 (1964) (White, J., dissenting) ("Without the evidence, the quest for truth may be seriously impeded and in many cases the trial court, although aware of proof showing defendant's guilt, must nevertheless release him because the crucial evidence is deemed inadmissible.").

383. Mathews v. Eldridge, 424 U.S. 319, 334-35 (1976); Morrissey v. Brewer, 408 U.S. 471, 481 (1972); Cafeteria & Rest. Workers Union v. McElroy, 367 U.S. 886, 895 (1961).

384. RICHARD H. FALLON, JR., THE DYNAMIC CONSTITUTION: AN INTRODUCTION TO AMERICAN CONSTITUTIONAL LAW 104 (2004); *see also* Solum, *supra* note 350, at 185-86 (describing procedural perfection as "unattainable" because of the unaffordable nature of "a system that achieved the highest possible degree of accuracy").

procedural framework adopted by the administrative agency."[385] The first of these is the somewhat rigid *Goldberg v. Kelly* approach, the second is the *Mathews* balancing test, and the third is in *Board of Curators v. Horowitz*,[386] where the Court denied a medical student's due process claim in connection with her expulsion from a state medical school.[387]

The *Mathews* balancing test is especially unsuitable for use in the war on terrorism, since the clash of individual and national security interests allows the judicial decisionmaker to reach any plausible outcome. The more deferential approach used in *Horowitz*, however, is even more unsuitable. Consider, for example, that the executive branch did not announce its Combatant Status Review Hearings until after the Court had decided in *Rasul v. Bush* that detainees held at Guantanamo Bay were entitled to petition for writs of habeas corpus.[388] Had the Court merely deferred to the Department of Defense's procedures, the Combatant Status Review Hearings may not have been implemented, and the Court would have indirectly upheld the informal procedures described by Secretary Rumsfeld.[389]

One possible alternative is a subset of the *Goldberg* approach, which Jerry Mashaw has labeled the "model of appropriateness," and involves using the analogical process to determine "appropriate" due process.[390] Although Mashaw sees the approach as promising "the maintenance of a moderately workable system,"[391] he also concludes that when "choices must be made among competing precedents and models, the appropriateness methodology provides no means for structuring the constitutional conversation."[392]

Mashaw may be correct in stating that the myriad due process cases (denial of discretionary benefits, academic expulsion or sanction, prison discipline, and loss of government employment) provide the courts with too many possibilities from which to choose with no guidance as to the relative appropriateness of the competing analogies.[393] But that criticism may have less force when focusing solely on the due process required in connection with involuntary, noncriminal detention of persons. Using those precedents, one can construct a baseline of procedural due process, with incremental additions or subtractions based on articulable distinctions between military detention and other forms of

385. William G. Buss, *Easy Cases Make Bad Law: Academic Expulsion and the Uncertain Law of Procedural Due Process*, 65 IOWA L. REV. 1, 49 (1979).

386. 435 U.S. 78 (1978).

387. *Id.* at 85-86 & n.2-3.

388. *See* Yin, *supra* note 57, at 1074.

389. *See supra* Part I.B.

390. MASHAW, *supra* note 348, at 53.

391. *Id.* at 97.

392. *Id.* at 99.

393. *See id.* at 99-100 (noting that *Goldberg v. Kelly* could be read either as a case about the government attempting to terminate a valid property interest without a hearing or as a case about one party being forced to pay money allegedly owed "to a judgment-proof claimant without the benefit of any security or bond" in order to ensure recovery of such money).

confinement.[394] Here, the single dimension of detention is more plausibly unifying among the various situations in which confinement occurs than in the other analogies.

In several cases, the Court has referred to the following elements as the "basic requirements" or "fundamental requirements" of due process: an "opportunity to be heard . . . at a meaningful time";[395] an unbiased decisionmaker;[396] "the right to present a defense" (including the right to testify and to call witnesses);[397] and adequate notice.[398] Two other frequently discussed elements of due process are the right to representation by counsel (including appointment of counsel for the indigent) and the right to confront and cross-examine. These two rights have been deemed sufficiently fundamental in the criminal prosecution context to have been included in the Sixth Amendment and made applicable to the States through incorporation via the Fourteenth Amendment's Due Process Clause. These components of due process (notice and hearing, unbiased decisionmaker, right to counsel, right to present a defense, and right to confront and cross-examine) will be analyzed under the proposed methodology.[399]

One potential objection to consider at the outset is the contention that the due process obligations that have evolved in the analogous situations to be studied were themselves the product of *Mathews* balancing, and therefore, one is doing little more than recycling *Mathews* balancing in a different guise. There is, admittedly, some truth to this objection. *Mathews* had a profound influence on procedural due process jurisprudence, and it cannot be denied that most of the cases concerning the due process rights of persons facing non-criminal detention have invoked *Mathews*. Indeed, *Mathews*' impact is such that we might question whether pre-*Mathews* decisions are still valid.[400] If *Mathews*, however, is not the appropriate approach for determining due process

394. *See* Rubin, *supra* note 337, at 1046 ("[M]odern procedural due process issues lend themselves to assessment by analogy.").

395. Armstrong v. Manzo, 380 U.S. 545, 552 (1965).

396. Withrow v. Larkin, 421 U.S. 35, 47 (1975); *see also* Weiss v. United States, 510 U.S. 163, 178 (1994) ("A necessary component of a fair trial is an impartial judge."); *In re* Murchison, 349 U.S. 133, 136 (1955) (requiring "the absence of actual bias" and attempting "to prevent even the probability of unfairness").

397. Washington v. Texas, 388 U.S. 14, 19 (1967).

398. Memphis Light, Gas, & Water Div. v. Craft, 436 U.S. 1, 22 (1978). These fundamental requirements of due process are also among the elements identified in Judge Friendly's influential article. Henry J. Friendly, *"Some Kind of Hearing,"* 123 U. PA. L. REV. 1267, 1279-87 (1975).

399. This hardly represents the full spectrum of due process-related matters that could arise in connection with the war on terrorism. The methodology suggested in this Article could also be applied to such elements as the burden of proof and the right to judicial review.

400. *Goldberg* remains good law, but that is because *Mathews* distinguished its facts from those of *Goldberg*. The generous approach to due process taken in *Goldberg* does not survive *Mathews* except for the specific holding regarding pretermination hearings for welfare recipients.

in the limited circumstances of the war on terrorism, that does not necessarily mean that its use in other analogous settings has produced incorrect outcomes such that those precedents should be disregarded.[401] These other precedents are used as a general baseline from which to determine whether similar procedures are constitutionally required in the special context of detention of individuals pursuant to congressionally authorized use of military force against non-state groups. This is the sort of analogical, common law model that is an important method of legal reasoning in American law.[402] This approach allows one to interpret the Due Process Clause with a degree of flexibility, so as to accommodate different circumstances, but also calls for the articulation of persuasive reasons for departing from the baseline norm. Therefore, if the Court has called for appointment of counsel in analogous situations, then the government must articulate plausible reasons why that the war on terrorism differs from those other situations, notwithstanding the apparent differences.

1. Notice and a Hearing

Once one establishes that a detainee is entitled to due process, it should be uncontroversial to conclude that there must be notice of the reason for the person's detention and a hearing in which the detainee is afforded the opportunity to challenge that detention. Even in cases such as *Mathews v. Eldridge*,[403] *City of Los Angeles v. David*,[404] *Connecticut v. Doehr*,[405] and *Mackey v. Montrym*,[406] the issue has not been whether the government may deny the claimant a hearing altogether. Rather, the central issue was whether the government must offer a pre-deprivation hearing, or whether a post-deprivation hearing sufficiently satisfied due process.[407] In other words, the default rule is that a hearing is required at some point before final deprivation.[408] Of course, when talking about deprivation of liberty, as opposed to property, "final deprivation" and "pre-" and "post-deprivation" are not

---

401. *But see* Buss, *supra* note 385, at 99-100 (arguing that in applying *Mathews* balancing, *Horowitz* may have reached the right result in upholding a medical school's expulsion of a student, but the wrong result by not requiring a hearing).

402. *See* FARBER & SHERRY, *supra* note 252, at 152-54 (discussing David Strauss, Cass Sunstein, Richard Posner, and Karl Llewellyn's support for analogical reasoning).

403. 424 U.S. 319 (1976).

404. 538 U.S. 715 (2003) (per curiam).

405. 501 U.S. 1 (1991). *But see* Bernard Schwartz, *A Decade of Administrative Law: 1987-1996*, 32 TULSA L.J. 493, 522 (1997) (arguing that the analysis used in *Doehr* could easily have resulted in the conclusion that no hearing of any sort was necessary).

406. 443 U.S. 1 (1979).

407. *See David*, 538 U.S. at 719 (holding that a 27-day delay in providing a hearing to challenge a parking infraction which resulted in a towed car did not violate due process); *Mackey*, 443 U.S. at 17, 19 (holding that the state's interest in "public health and safety" justified suspension of a driver's license prior to a hearing).

408. *Mathews*, 424 U.S. at 349.

especially useful concepts because the deprivation continues for as long as the person is involuntarily detained.

What does due process require as the focus of the hearing? Because the action taken against an enemy combatant is the deprivation of liberty, and because the point of due process is to minimize the likelihood of erroneous deprivation of that liberty, one should consider the source of the President's power to detain. The President has the Authorization to Use Military Force ("AUMF") against "those nations, organizations, or persons [the President] determines planned, authorized, committed, or aided the terrorist attacks that occurred on September 11, 2001, or harbored such organizations or persons."[409] Curtis Bradley and Jack Goldsmith argue that this joint resolution covers those "persons who receive and execute orders within [Al Qaeda's] command structure,"[410] as well as those "individuals who take up arms for purposes of attacking the United States on [Al Qaeda's] behalf, and also those who are in the process of 'prepar[ing] for combat and return[ing] from combat.'"[411] Therefore, assuming that the joint resolution provides the President with the substantive power to detain persons who fall within the definition of "enemy,"[412] membership in Al Qaeda would justify detention of the captured person.[413] There are, of course, some difficult questions at the margins, such as whether the AUMF would authorize detention of a person who belonged to a terrorist organization that had no connection to the 9/11 attacks and did not become affiliated with Al Qaeda until after September 11, 2001.[414] But the purpose of the hearing is whether the detainee is someone who falls within the scope of the AUMF—not simply with potential criminal conduct.

2. Unbiased Decisionmaker

It should also be uncontroversial to conclude that due process requires that the hearing officers be neutral and unbiased because the Court has emphasized that a "biased decisionmaker" is "constitutionally unacceptable."[415] In numerous contexts, the Court has rejected the presence of a biased decisionmaker.[416] Thus, there is nothing exceptional about *Hamdi*'s holding

---

409. Authorization for Use of Military Force, *supra* note 2.

410. Bradley & Goldsmith, *supra* note 2, at 2114-15.

411. *Id.* at 2116 (quoting Int'l Comm. of the Red Cross, Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949 ¶ 1943 (1987)).

412. *See id.* at 2066-72 (discussing whether Congress can authorize military force against a non-state actor group); Yin, *supra* note 19, at 189-90.

413. Or the Taliban.

414. *See* Bradley & Goldsmith, *supra* note 2, at 2109-10.

415. Withrow v. Larkin, 421 U.S. 35, 47 (1975).

416. *See* Taylor v. Hayes, 418 U.S. 488, 501-03 (1974) (barring the judge from hearing a case on remand where the judge "became personally involved" in proceedings as evidenced by biased trial conduct); Gibson v. Berryhill, 411 U.S. 564, 578-79 (1973) (holding that State Board of Optometry was biased by pecuniary self interest and thereby disqualified from making

that an American citizen captured on the battlefield is entitled to a hearing before a "neutral decisionmaker."

While it is clear that due process requires an unbiased decisionmaker, the more challenging due process inquiry is what constitutes "bias" beyond the obvious examples of self-interest (financial or otherwise) and personal involvement in the matter being adjudicated. If the decisionmaker is a military officer who played no role in the capture or initial classification of the detainee as an enemy combatant, is the officer, nevertheless, "biased" simply because as an executive branch agent, the officer is ultimately beholden to the President and likely to rule in favor of the government even when such a decision is not objectively warranted?[417] The debate on this issue tends to focus on whether the decisionmaker should be an elected federal judge, a civilian judge, or a military judge.[418] Those who support federal judges as the decisionmakers argue that only judges with life tenure can be trusted to not succumb to political or public pressure. The Court, however, has yet to hold that due process requires the decisionmaker to have life tenure. Perhaps the Court has not limited appropriate decisioinmakers to those with life tenure because many state court judges are elected by the public. Such judges preside over the vast majority of criminal prosecutions in the United States without running afoul of the Fourteenth Amendment's Due Process Clause.[419]

---

decisions); Ward v. Village of Monroeville, 409 U.S. 57, 59, 61-62 (1972) (holding that a mayor authorized by statute to be a judge was not "neutral and detached" where part of his salary was derived from court fees); Mayberry v. Pennsylvania, 400 U.S. 455, 466 (1971) (barring a judge from hearing the case on remand where the lawyer had personally insulted the judge several times, resulting in a contempt order); Pickering v. Bd. of Educ., 391 U.S. 563, 578 n.2 (1968) (noting the potential bias of a school board in the position of deciding the consequences of a teacher's attack on the same board); Tumey v. Ohio, 273 U.S. 510, 523 (1927) (holding that defendant's due process rights had been violated where the judge had a "direct, personal, substantial, pecuniary interest" in the case); *cf.* Commonwealth Coatings Corp. v. Cont'l Cas. Co., 393 U.S. 145, 149 (1968) (requiring arbitrators to "disclose to the parties any dealings that might create an impression of possible bias"); Ungar v. Sarafite, 376 U.S. 575, 584 (1964) (finding no bias where the judge's contempt ruling was occasioned by a lawyer's refusal to answer questions). *But see* Hortonville Joint Sch. Dist. No. 1 v. Hortonville Educ. Ass'n, 426 U.S. 482, 491-94 (1976) (concluding that school board members had no personal stake in the decision to fire striking teachers and hence were not biased decisionmakers); Arnett v. Kennedy, 416 U.S. 134, 155 n.21 (1974) (allowing a pretermination of employment hearing to be headed by the very person who was victimized by the employee).

417. *See* THE FEDERALIST NO. 78, at 471 (Alexander Hamilton) (Willmoore Kendall & George W. Carey eds., 1966) ("Periodical appointments, however regulated, or by whomsoever made, would, in some way or other, be fatal to their necessary independence.").

418. Yin, *supra* note 57, at 1076-78 (summarizing relevant Supreme Court cases).

419. *See* Republican Party of Minn. v. White, 536 U.S. 765, 782-83 (2002) (observing that the mere fact that elected judges experience electoral pressures does not automatically trigger due process concerns); *see also* Weiss v. United States, 510 U.S. 163, 179, 181 (1994) (rejecting a due process challenge to composition of courts-martial based on the lack of fixed terms or life tenure of military judges).

In addition to not requiring life tenure, due process may not even require that the decisionmaker be legally trained, depending on the nature of the liberty deprivation.[420] Consistent with due process, a trained medical professional can serve as the decisionmaker when the issue is whether to administer medication to mentally ill prisoners by force,[421] to transfer a prisoner to a mental hospital,[422] or "so long as [the medical professional] is free to evaluate independently,"[423] to admit a juvenile to a mental hospital.[424]

In the medical decisionmaker cases, the Court was no doubt swayed by the fact that the issues were fundamentally medical and scientific in nature, meaning an "untrained judge or administrative hearing officer" would not be expected to reach a more accurate decision than a medical professional.[425] The war on terrorism, however, is different because the ultimate issue of whether a given detainee is an enemy combatant is not one that requires a "medical-psychiatric" conclusion.[426] Moreover, a special relationship exists between medical professionals and their patients because the former must act in the best interests of the latter.[427] Soldiers, on the other hand, bear no such relationship to their prisoners, and in fact, the obligation that soldiers owe is to protect their nation from the enemy. While the law of armed conflict does require a soldier to take reasonable steps to *minimize* civilian and noncombatant casualties (suggesting that soldiers may not inflict violence indiscriminately), it does not *forbid* so-called collateral damage.[428]

Even if the presence of a special relationship, such as the physician-patient relationship, might strengthen the reasonableness of allowing nonlegally trained decisionmakers, the absence of such a relationship need not compel the opposite conclusion. For example, the prison discipline cases demonstrate that nonlegally trained prison officials can serve as the decisionmakers. As captors, prison officials have a constitutional obligation to protect such constitutional rights as due process.[429] This constitutional obligation, however, does not

---

420. Parham v. J.R., 442 U.S. 584, 607 (1979) ("Due process has never been thought to require that the neutral and detached trier of fact be law trained or a judicial or administrative officer.") (citing Morrisey v. Brewer, 408 U.S. 471, 489 (1972); Goldberg v. Kelly, 397 U.S. 254, 271 (1970).

421. Washington v. Harper, 494 U.S. 210, 227, 231 (1990).

422. Vitek v. Jones, 445 U.S. 480, 496 (1980).

423. *Parham*, 442 U.S. at 607.

424. *Id.* at 606-07.

425. *Id.* at 609.

426. *But see* Yin, *supra* note 57, at 1084-85 (exploring reasons why "enemy combatant" determination requires military expertise).

427. *See, e.g.*, Jennifer Russano, *Is Boutique Medicine a New Threat to American Health Care or a Logical Way of Revitalizing the Doctor-Patient Relationship?*, 17 WASH. U. J.L. & POL'Y 313, 330 & n.11 (2005).

428. *See, e.g.*, Shannon E. French, *Murderers, Not Warriors*, *in* TERRORISM AND INTERNATIONAL JUSTICE 31, 38-39 (James P. Sterba ed., 2003) (discussing the Doctrine of Double Effect and unintentional harm to innocents during warfare).

429. Wolff v. McDonnell, 418 U.S. 539, 555-56 (1974) (noting that prisoners enjoy several

create a special relationship, and in any event, is an obligation largely owed by the military to its captives as well.

In short, the confinement cases do not require the prophylactic implementation of life tenure or fixed terms of office, nor do they require that decisionmakers be legally trained in order to ensure lack of bias.

3. Appointment of Counsel for the Indigent

Another key issue is whether a detainee has a due process right to be represented by counsel, including the appointment of counsel in the case of an indigent detainee. Any right to counsel would arise from the Fifth Amendment's Due Process Clause, not the Sixth Amendment, because the Sixth Amendment right to counsel applies only in criminal prosecutions,[430] whereas the persons detained at Guantanamo Bay are not being prosecuted but instead, are being detained under the law of armed conflict.[431] Because of the nature of a detainee's detention, the landmark case of *Gideon v. Wainwright*,[432] which held that the Sixth Amendment obligated the states to provide counsel for indigent defendants facing felony charges, does not directly apply to this noncriminal situation.

The first case to find a due process right to counsel in a noncriminal setting was *In re Gault*,[433] which held that an indigent juvenile was entitled to the appointment of counsel because he faced the potential loss of personal freedom.[434] Although the Court acknowledged that the juvenile in this case was not classified as a "criminal,"[435] the Court also recognized that the nature of the detention was more important:

> A boy is charged with misconduct. The boy is committed to an institution where he may be restrained of liberty for years. It is of no constitutional consequence . . . that the institution to which he is committed is called an Industrial School. The fact of the matter is that, however euphemistic the title, a "receiving home" or an "industrial school" for juveniles is an

---

constitutional protections, including, *inter alia*, rights under the Due Process Clause).

430. *See* Lassiter v. Dep't of Soc. Servs., 452 U.S. 18, 25 (1981) (observing precedent that Sixth and Fourteenth Amendments' right to counsel applies "in criminal cases"); *see also* Middendorf v. Henry, 425 U.S. 25, 38 (1976) (noting that "even in the civilian community a proceeding which may result in deprivation of liberty is nonetheless not a 'criminal proceeding' within the meaning of the Sixth Amendment if there are elements about it which sufficiently distinguish it from a traditional civilian criminal trial").

431. *See* Yin, *supra* note 19, at 152-53 (pointing out the post-9/11 debate over the use of criminal law versus the law of armed conflict).

432. 372 U.S. 335 (1963); *see also* Argersinger v. Hamlin, 407 U.S. 25, 37 (1972) (extending *Gideon* to require appointment of counsel, absent waiver, even in cases involving petty crimes or misdemeanors so long as the defendant faces potential imprisonment).

433. 387 U.S. 1 (1967).

434. *Id.* at 41.

435. *Id.* at 23.

institution of confinement in which the child is incarcerated for a greater or lesser time.[436]

Gault's principle of looking at the true nature of the detention has expanded the due process requirement of the appointment of counsel to other instances of noncriminal detention. While the Court has not considered whether due process requires the appointment of counsel to represent indigents in civil commitment or civil contempt proceedings, lower courts in various jurisdictions had already reached that conclusion by using *Gault*'s deprivation of physical liberty standard.[437] It is also significant that virtually all states statutorily provide indigents with appointed counsel in civil commitment proceedings,[438] indicating that there is a well-established norm of protecting against erroneous deprivations of liberty.

Therefore, modern practices amply support the Court's recognition of a "presumption that an indigent litigant has a right to appointed counsel only when, if he loses, he may be deprived of his physical liberty."[439] Of course, there are instances where a person faces loss of physical liberty yet is not constitutionally entitled to appointment of counsel. In *Gagnon v. Scarpelli*,[440] the Court held that the appointment of counsel to represent indigent defendants facing revocation of probation could be determined on a case-by-case basis by the trial court.[441] In *Wolff v. McDonnell*,[442] the Court held that prison inmates facing internal disciplinary proceedings were not constitutionally entitled to counsel.[443] In *Vitek v. Jones*,[444] only four Justices were willing to hold that an indigent prisoner facing involuntary transfer to a mental hospital was entitled to appointment of counsel;[445] the deciding vote in the case came from Justice

---

436. *Id.* at 27.

437. Heryford v. Parker, 396 F.2d 393, 396 (10th Cir. 1968); Dorsey v. Solomon, 435 F. Supp. 725, 733 (D. Md. 1977); Lessard v. Schmidt, 349 F. Supp. 1078, 1098 (E.D. Wis. 1972); *see also* Specht v. Patterson, 386 U.S. 605, 610 (1967) (holding that Colorado sentencing statute was of such a nature that due process afforded defendant the right to counsel); Robert S. Catz & Nancy Lee Firak, *The Right to Appointed Counsel in Quasi-Criminal Cases: Towards an Effective Assistance of Counsel Standard*, 19 HARV. C.R.-C.L. L. REV. 397, 412-15 (1984) (discussing right to appointed counsel in civil contempt cases).

438. *See* Catz & Firak, *supra* note 437, at 411; *see also* Kansas v. Hendricks, 521 U.S. 346, 353 (1997) (citing with approval a Kansas statute providing appointment of counsel for indigent sex offenders facing civil commitment); *Specht*, 386 U.S. at 610 (holding that a convicted sex offender facing commitment under a civil statute had the due process right to appointed counsel).

439. Lassiter v. Dep't of Soc. Servs., 452 U.S. 18, 26-27 (1981).

440. 411 U.S. 778 (1973).

441. *Id.* at 790.

442. 418 U.S. 539 (1974).

443. *Id.* at 570; *see also* Baxter v. Palmigiano, 425 U.S. 308, 315 (1976) (affirming the *Wolff* holding that prisoners have no right to counsel in disciplinary proceedings).

444. 445 U.S. 480 (1980).

445. *Id.* at 496-97 (White, J., joined by Brennan, Marshall & Stevens, JJ.).

Powell, who concluded that a competent layperson could provide the necessary assistance to the prisoner.[446]

Regardless of whether these cases were correctly decided, one important element they share is that they involved detainees who were convicted of crimes and were either facing a change in detention conditions or, in the case of *Gagnon*, a return to prison while still serving a criminal sentence. Convicted prisoners, however, are obviously entitled to no due process, but the deprivation of liberty involved is of much lesser degree than it would be for a person who has not been convicted of a crime and faces confinement for the same length of time and under the same conditions. Even the probationer in *Gagnon* had a lesser liberty interest at stake because the terms of probation typically include mildly onerous conditions, such as frequent meetings with probation officers, drug testing, travel restrictions, and what amounts to forfeiture of Fourth Amendment privacy rights.[447]

Because due process has required appointment of counsel in other noncriminal detention proceedings, however, does not mean that the same must be true in the war on terrorism. It does suggest, however, that if one deviates from the norm established in similar contexts, there should be plausible reasons explaining why due process would require something different in this setting. The most immediate difference is that the war on terrorism implicates national security interests, and as previously stated, "national security" is often, though not always, a magic phrase that empowers the government to do that which it would otherwise be prohibited from doing.[448]

Yet, no one has seriously suggested that federal criminal defendants in prosecutions implicating national security interests, such as ones involving the potential disclosure of classified information or espionage, can be stripped of their right to counsel, or even that their rights are different at all because of the presence of national security interests. Similarly, in civil cases implicating national security interests, the government does not argue that litigants can be deprived of their private counsel. The most that can be said is that the government's national security interest may justify the imposition of prophylactic measures to ensure that the attorney-client relationship does not harm national security.[449] Even those prophylactic measures are hardly as radical as they might seem. It has been long-established that the attorney-client privilege may not be used to help further future crimes and that evidence of such intent may justify piercing the privilege and forcing the disclosure of

446. *Id.* at 500 (Powell, J., concurring in part).

447. *See, e.g.*, Griffin v. Wisconsin, 483 U.S. 868, 874 (1987) (observing that federal statutes place a variety of restrictions on probationers).

448. *See supra* note 365.

449. For example, the Department of Justice presently authorizes "the monitoring or review of communications between . . . inmate[s] and attorneys" in exigent circumstances. 28 C.F.R. § 501.3(d) (2005).

attorney-client communications.[450] National security measures merely make it easier for the government to identify and prosecute such transgressors.[451]

In short, there is no precedent, custom, or practice that would suggest that due process requires the appointment of counsel for indigent litigants who face potential loss of physical freedom in all cases, except for ones involving national security. Thus, one might substitute "war" for "national security," and consider whether the result would be any different. War is obviously different in that it empowers soldiers to do things that they would otherwise be prohibited from doing, such as killing other persons, whether in self-defense or not.[452] Likewise, there is no historical precedent for the appointment of counsel to enemy prisoners of war in order to challenge the legality of their confinement. The lack of precedent can be fully explained by the enemy alien disability rule, which prohibits enemy aliens from accessing U.S. courts during war time to hamper the war efforts.

Moreover, in a traditional war, the number of enemy prisoners of war might be staggering; at the end of World War II, the United States held as many as two million POWs.[453] Providing counsel to that many persons on an individual basis would be impossible. If that point is persuasive, however, one must consider that the number of worldwide Al Qaeda members is much less.[454]

The final difference between the war on terrorism and other detention contexts is the government's asserted need to interrogate the Guantanamo Bay detainees for counterterrorism purposes.[455] Some argue that given the catastrophic scope of the 9/11 attacks and given Al Qaeda's desire to inflict even worse damage in subsequent terrorist attacks on the United States, it is imperative to extract whatever intelligence possible to help the executive branch avert such further attacks.[456] But providing counsel to detainees would

---

450. *See, e.g., In re* Sealed Case, 107 F.3d 46, 49 (D.C. Cir. 1997) (explaining the "crime-fraud exception" to the attorney-client privilege).

451. In one notable case, a well-known criminal defense attorney, Lynne Stewart, was convicted of numerous charges based on evidence developed as a result of government monitoring of attorney-client communications. *See* Julia Preston, *For Lawyer's Lawyer, Loss Is Just the Beginning*, N.Y. TIMES, Feb. 18, 2005, at B4; *see generally* United States v. Sattar, 314 F. Supp. 2d 279 (S.D.N.Y. 2004) (denying Lynne Stewart's motions to dismiss).

452. For example, the law of armed conflict permits ambushes and other ruses of war, even though in an ambush the target obviously presents no immediate threat to the ambusher.

453. *See* Rasul v. Bush, 542 U.S. 466, 498 (2004) (Scalia, J., dissenting) (citing GEORGE G. LEWIS, HISTORY OF PRISONER OF WAR UTILIZATION BY THE UNITED STATES ARMY 1776-1945 244 (1955)).

454. One study estimates that since 1996 approximately 20,000 men passed through Al Qaeda training camps in Afghanistan. Of those, it is estimated that 2,000 were killed or captured by the United States. INT'L INST. FOR STRATEGIC STUDIES, STRATEGIC SURVEY 2003/4 6 (Jonathan Stevenson ed., 2004).

455. *See* Woolfolk Declaration, *supra* note 33, at 554.

456. *See* Lee A. Casey & David B. Rivkin, Jr., *How to Treat A Captured Terrorist: Getting to the Heart of an Important Question*, NAT'L REV., July 4, 2005, at 20 (arguing that applying the Geneva Conventions to al-Qaeda detainees would hamper intelligence gathering

undermine the effectiveness of interrogators, who need to isolate their targets and build rapport and dependency in order to interrogate effectively. The argument is not without force and may well justify a period of delay before providing counsel, but it would hardly support a blanket denial of counsel from now until the end of the war on terrorism. At some point, whatever information a detainee has relevant to the war on terrorism will become stale and useless. Moreover, the longer the period of detention without access to counsel for a hearing, the greater the degree of deprivation suffered by the detainee.[457]

The result under a *Mathews* balancing test could quite plausibly be different. One could argue that the marginal gain in accuracy from having a detainee represented by counsel would be slight because the factual issue to be resolved—whether a detainee fits within the category of persons against whom Congress has authorized the use of military force—might not involve complex legal issues.[458] Just as *Vitek* held that a prisoner facing involuntary transfer to a mental hospital could be ably assisted by a competent medical professional because the issue was medical, not legal,[459] a court analyzing the increase in accuracy because of counsel with regard to detainees in the war on terrorism could conclude that a military officer could ably assist a detainee.[460]

Indeed, one could even take the radical view that the provision of counsel may lead to a *decrease* in the accuracy of the results.[461] The traditional view is that "accurate and just results are most likely to be obtained through the equal contest of opposed interests" including counsel, "without whom the contest of interests may become unwholesomely unequal."[462] That much is not in doubt, but "just results" may not be the same as "accurate results" in an adversarial system. Consider, for example, a criminal case in which the defense lawyer moves successfully to suppress critical evidence against her client on the basis that it was obtained in violation of the Fourth Amendment. This is a "just" result in the sense that the need to enforce the Fourth Amendment via the exclusionary rule outweighs any individual case; however, it is hardly an "accurate" result, for the defendant (absent other defense) may be clearly guilty

---

capabilities); John Hendren, *Rumsfeld Okayed Harsher Methods of Interrogation*, STAR-LEDGER, May 21, 2004, at 4 (noting government defense of harsh interrogation tactics).

457. *Cf.* Fusari v. Steinberg, 419 U.S. 379, 389 (1975) ("[T]he possible length of wrongful deprivation of . . . benefits is an important factor in assessing the impact of official action on the private interests").

458. Yin, *supra* note 57, at 1085-88 (exploring the pros and cons of providing detainees with counsel).

459. Vitek v. Jones, 445 U.S. 480, 495-96 (1980).

460. Yin, *supra* note 57, at 1084-85 ("[A]ny combat soldier . . . should be able to help a detainee address allegations of being an enemy combatant. American soldiers receive training on rules of engagement, which are 'a condensed version of the laws of war.'" (quoting Paul Maliszewski & Hadley Ross, *We Happy Few: The U.S. Soldier's Laws of War, in Principle and in Practice*, HARPER'S MAG., May 2003, at 56)).

461. To be clear, this argument is set forth not to endorse it but to highlight the problems with using the *Mathews* balancing test in determining due process in the war on terrorism.

462. Lassiter v. Dep't of Soc. Servs., 452 U.S. 18, 28 (1981).

from a factual standpoint. Also consider the classic ethical question faced by a criminal defense attorney who knows that her client is factually guilty. How may she proceed to defend her client? She is certainly entitled to hold the government to its burden of proving guilt beyond a reasonable doubt,[463] which may include cross-examining prosecution witnesses in an effort to limit the persuasiveness of their testimony. Acquittal of the defendant might be just, but not accurate, given the defendant's literal guilt. Therefore, the issue of entitlement to appointment of counsel is one in which the mode of due process analysis suggested herein could well lead to a different result than one might reach under a *Mathews* balancing test.

4. Right to Present a Defense

A hearing in which the detainee has no right to present a defense is little more than a forum for explaining to the detainee why the United States is holding him against his will. While the detainee might derive some satisfaction in getting an explanation, that satisfaction is likely to result in overwhelming frustration if it turns out that the government's underlying assumptions are erroneous. This explains why the Court has viewed the right to present a defense, including calling witnesses and testifying, as a "basic" part of due process.[464]

Yet, the right to call witnesses is not absolute. In the prison discipline cases, for example, the Court has consistently upheld prison officials' authority to deny a prisoner facing a disciplinary hearing the right to call witnesses where "granting the request would be 'unduly hazardous to institutional safety or correctional goals.'"[465] Some of the reasons for denying prison inmates the right to call witnesses may seem applicable to doing so in hearings for terrorism detainees; other reasons may seem inapplicable. "[T]he penological need to provide swift discipline,"[466] for example, is irrelevant to detentions under military authority. There may be a need in the heat of battle to isolate potential enemy fighters quickly and to remove them from the battlefield without the need for cumbersome hearings, but once the detainees have been moved to a secure location, the emergency rationale for denying hearings altogether is no longer valid. Furthermore, detention pursuant to the law of armed conflict has

---

463. *See* MODEL RULES OF PROF'L CONDUCT R. 3.1 (2002) (providing that counsel for the dependant "may . . . so defend the proceeding as to require that every element of the case be established").

464. *See, e.g.*, Ponte v. Real, 471 U.S. 491, 495 (1985) (noting that prison inmates have due process rights "to call and present witnesses and documentary evidence in [their] defense"); Washington v. Texas, 388 U.S. 14, 18-19 (1967) (observing previous holdings that due process affords a defendant the opportunity to call and confront witnesses as well as testify in criminal proceedings).

465. *Ponte*, 471 U.S. at 495 (quoting Wolff v. McDonnell, 418 U.S. 539, 566 (1974)).

466. *Id.*

no penological goal; detention is justified only as a matter of preventative incapacitation.[467]

On the other hand, the need to avoid exposing detention facilities to unduly hazardous situations may be of some relevance to our situation. A detainee who requests that dozens of Taliban fighters being held in Afghanistan attend his hearing might be increasing the prisoner-guard ratio at Guantanamo Bay to an intolerably high level. Where the number of witnesses is relatively small, however, this security concern might not justify denial of the right to call those witnesses. As with the prison discipline cases, one might expect courts to show some deference, though not outright abdication, to government officials' threat assessment.

There are likely to be practical obstacles, including some that are severe, to fulfilling detainees' desires to call witnesses. Some witnesses may be citizens of countries hostile to the United States,[468] others may be difficult or impossible to locate, others may be dead, and still others may be in military custody but unavailable for reasons of national security. The last reason is no undoubtedly the most controversial. If one accepts the premise that national security can justify making otherwise relevant witnesses unavailable for a hearing on the legality of detention, then the likely outcome will be deference to the executive branch's determination that making the witness available will, in fact, harm national security.

As it turns out, the criminal justice system has faced this problem already. During pre-trial discovery, accused 9/11 conspirator Zacarias Moussaoui wanted to interview Al Qaeda operatives Ramzi Binalshibh and Khalid Sheikh Mohammed, who were both in U.S. custody in undisclosed locations. Although Moussaoui did not deny being a member of Al Qaeda, he disputed the government's contention that he was to have been the "20th hijacker"; according to Moussaoui, he was to have been part of a post-9/11 terror plot but

---

467. This refers to the decision to detain a given person, where the hearing focuses on whether the person has been identified correctly as subject to detention, and not the decision to punish a detainee for violations of the laws of war, or, if the Geneva Convention were deemed applicable, for violations of the rules governing the conduct of prisoners of war. *See* Geneva Convention, *supra* note 7, arts. 89-90 (detailing duration and types of appropriate punishments for disciplinary infractions by prisoners of war).

468. Practically speaking, a country might be considered hostile despite having a friendly government if the local population evinces anti-American sentiments. Pakistan, for example, is ruled by General Musharaff, who is considered an ally of the United States. However, polls show that Pakistanis have a largely negative view of the United States, especially compared to Osama bin Laden. *See, e.g.*, PEW RESEARCH CTR., PEW GLOBAL ATTITUDES PROJECT, U.S. IMAGE UP SLIGHTLY, BUT STILL NEGATIVE (2005), *available at* http://pewglobal.org/reports/display.php?ReportID=247; *see also* PEW RESEARCH CTR., PEW GLOBAL ATTITUDES PROJECT, SPRING 2005 17-NATION SURVEY 45-46 (2005), *available at* http://pewglobal.org/reports/pdf/248topline.pdf (noting 29% of Pakistani respondents reporting "a lot of confidence" and 22% reporting "some confidence" in bin Laden "to do the right thing regarding world affairs"). One can imagine that it might be difficult to persuade a Pakistani citizen to come to Guantanamo Bay to testify at a hearing involving a detainee.

not part of the 9/11 plot itself, a contention that he believed Mohammed and Binalshibh would support. The government argued that providing Moussaoui with access of any kind to Mohammed and Binalshibh would interfere with ongoing interrogation of the two Al Qaeda members and would therefore endanger national security.[469] Instead, the government offered to provide Moussaoui with summaries of relevant responses by Mohammed and Binalshibh to interrogation.

At first glance, the government actions seem like a clear violation of *Brady v. Maryland*,[470] which held that it is a due process violation when the prosecution fails to hand over to the defense any exculpatory evidence it has.[471] In Moussaoui's case, however, the government was able to plausibly argue that it did not have the exculpatory evidence sought by Moussaoui because it had never asked Mohammed and Binalshibh the questions that Moussaoui wanted to ask. But the government still has custody of the persons who could answer those questions, if those question had only been posed. Many jurisdictions require the prosecution to make documents or other tangible physical evidence available to the defense so that the defense can conduct its examination.[472]

In Moussaoui's case, the district judge ordered the government to make Mohammed and Binalshibh available to Moussaoui for questioning, and when the government refused, the judge sanctioned the government by eliminating the death penalty as a possible sentence in the event of conviction and also by forbidding the government from introducing any evidence related to the 9/11 attacks. On appeal, the Fourth Circuit agreed that Binalshibh and Mohammed had material testimony to offer, that Moussaoui was entitled to their testimony, but that the appropriate sanction was to require the government to work with Moussaoui and the district court to produce acceptable substitutions to present to the jury in lieu of the actual witnesses.[473]

The upshot is that if the government need not produce high-level captives for a criminal trial, where the defendant has Sixth Amendment rights, it is difficult to see how the Due Process Clause would require a different outcome in noncriminal proceedings. Yet the reasons for denying prison inmates the right to call witnesses for disciplinary hearings are less compelling when applied to the Guantanamo detainees, and certainly friendly witnesses in Afghanistan or Pakistan, for example, could possibly appear via video teleconferencing.[474]

---

469. *See* United States v. Moussaoui, 382 F.3d 453, 470 (4th Cir. 2004).

470. 373 U.S. 83 (1963).

471. *Id.* at 87; *see also* Strickler v. Greene, 527 U.S. 263 (1999); Kyles v. Whitley, 514 U.S. 419 (1995).

472. *See, e.g.*, FED. R. CRIM. P. 16(a)(1)(E) (directing the government to allow the defendant, upon request and under specific circumstances, access to evidence in the government's control).

473. *Moussaoui*, 382 F.3d at 479-80.

474. *See* Neuman, *supra* note 115, at 61-62 (noting that "modern habeas procedure dispenses with production of the body in situations where the petitioner's claims can be

IV. CONCLUSION

If due process is fundamentally about freedom from arbitrary government action, then it has an especially important role to play in the war on terrorism, in which the government claims the power to detain uncharged persons in the name of "national security."[475]

Unfortunately, neither *Rasul v. Bush*[476] nor *Hamdan v. Rumsfeld*[477] answered the key question of whether nonresident aliens, such as the Guantanamo detainees, have any due process rights. This article argues that the answer is that they do have a cognizable liberty interest in being free from confinement. The response to those who fear that such recognition would inevitably lead to an ocean of habeas petitions brought by enemy soldiers challenging their confinement during a full-scale armed conflict is to invoke the well-established enemy alien disability rule. But the enemy alien disability rule is particularly ill-suited for use in the war on terrorism, where there is no nation-state "enemy" and where the detainees' membership in the authorized target of military force is itself the disputed issue.

In determining the scope of due process available to the detainees, this Article suggests that the *Mathews* balancing test is problematic because it involves balancing two absolute interests with no guidance on how to conduct that balancing, and because many of the procedures that the government might use could even lead to a decrease in accuracy. Instead, this Article argues for use of an analogical reasoning model, which suggests that the Guantanamo detainees should be afforded a hearing, notice, appointed counsel, and a reasonable right to present evidence, including possible teleconferencing of witnesses from abroad. Reasonable minds could, of course, disagree about the outcome of the particular procedures analyzed within this Article; perhaps persuasive arguments could be made as to why, for example, counsel should not be appointed to represent detainees in hearings regarding their status as enemy combatants. But freed from *Mathews*' straitjacket of utilitarianism and accuracy in a world where such determinations are virtually impossible to make, such counter-arguments would stand or fall on their own persuasiveness, instead of being cloaked within the language of cost-benefit analysis.

---

resolved as a matter of law without testimony").

475. *Cf.* Schwartz, *supra* note 405, at 524 ("[T]he courts must be vigilant in ensuring that flexible due process does not result in dilution of due process.").

476. 542 U.S. 466 (2004).

477. 126 S. Ct. 2749 (2006).

ENRON-INSPIRED NONQUALIFIED DEFERRED COMPENSATION RULES: "IF YOU DON'T KNOW WHERE YOU'RE GOING, YOU MIGHT NOT GET THERE."[1]

WILLIAM A. DRENNAN*

Enron's top 200 executives averaged $7.45 million each in compensation the year before Enron went bankrupt.[2] Former Tyco CEO, Dennis Kozlowski, received a $600 million compensation package that included the use of an $18 million Manhattan apartment, $30,000 opera glasses, a $16,000 dog-shaped umbrella stand, a $6,000 shower curtain in the maid's bathroom, and a $2 million toga party on the Mediterranean island of Sardinia featuring singer Jimmy Buffett and X-rated ice sculptures—all this, even though Tyco shareholders lost $80 billion during his reign.[3] But rather than addressing the fundamental causes of excessive executive compensation at public corporations, Congress enacted an Internal Revenue Code ("I.R.C.") provision that merely imposes timing rules on all employers and employees. For example, Kenneth Lay, ex-Enron CEO, deferred $32 million into a nonqualified deferred compensation ("NQDC") plan in one year.[4] In response to such abuse, Congress enacted section 409A of the Internal Revenue Code

---

1. YOGI BERRA WITH DAVE KAPLAN, WHEN YOU COME TO A FORK IN THE ROAD, TAKE IT! 53-54 (Hyperion 2001) ("[Y]ou can get to where you want to go if you know where you're going. What I mean is you must have a good idea of what you want. You need a plan.").

* Assistant Professor, Southern Illinois University School of Law; Member of the American Law Institute; co-author of the American Bar Association's book on Nonqualified Deferred Compensation; LL.M. in Intellectual Property, Washington University (2003); LL.M. in Taxation, Washington University (1997); J.D., St. Louis University (1985). The author would like to thank Eileen Schwartz for her enthusiastic assistance with the manuscripts, and Professor Michael Hussey for his comments.

2. In 2000, Enron's 200 most highly compensated executives received $1.06 *billion* from stock options, $172.6 million in wages, $131.7 million in restricted stock, $56.6 million in bonuses, and deferred compensation of $67 million, for a total of $1.49 billion. Enron declared bankruptcy on December 2, 2001. *See* JOINT COMMITTEE ON TAXATION, 108TH CONG., REPORT OF INVESTIGATION OF ENRON CORPORATION AND RELATED ENTITIES REGARDING FEDERAL TAX AND COMPENSATION ISSUES, AND POLICY RECOMMENDATIONS 13-14, 603 (JCS-3-03) (February 2003), *available at* http://www.gpo.gov/congress/joint/jcs-3-03/vol1/index.html [hereinafter Enron Compensation Report]; *see also infra* notes 90 and 94 (detailing the compensation for the 200 highest paid Enron executives for the years 1998 to 2000 and the amount of compensation those top 200 deferred from 1998 to 2000).

3. *See infra* note 26.

4. Enron Compensation Report, *supra* note 2, at 14. Although Mr. Lay was convicted in May of 2006 for deceiving investors about Enron's finances, the conviction was vacated on October 17, 2006 because Mr. Lay died on July 5, 2006). Greg Farrell, *Trial Judge Vacates Conviction of Late Enron Founder Lay*, USA TODAY, Oct. 18, 2006, at 3B.

("Section 409A"). Under this new law, a provision that seeks to prevent a CEO from deferring millions of dollars may also prevent a parent who is experiencing financial difficulties from withdrawing a few thousand dollars from her NQDC plan to help pay her child's tuition, even if her employer consents.[5]

I. INTRODUCTION

With financial scandals involving Kenneth Lay at Enron, Bernie Ebbers at WorldCom, Dick Grasso at the New York Stock Exchange, and Dennis Kozlowski at Tyco, just to name a few, highlighting the opening of the 21st Century, we begin with a multiple-choice question:

> What is the most serious problem with executive compensation today?
>
> a. The payment of grossly excessive compensation to top executives at publicly-traded corporations.[6]
> b. The failure of boards of directors to bargain at arm's-length with top executives at public companies regarding compensation, and the absence of structures and effective incentives to encourage arm's-length bargaining.[7]
> c. The compensation of top executives at public companies is not adequately disclosed to shareholders, prospective investors, and others.[8]
> d. Middle- and lower-management employees are taxed at roughly the same marginal income tax rates as top executives earning tens of millions of dollars.[9]
> e. The ability of an employer and an employee to agree that a portion of the employee's compensation will be deferred and paid in a future tax year at a specified time or upon the occurrence of a specified event, and to subsequently change the date or the event that will trigger payment to the employee.[10]

If you chose "e," you likely were delighted with the enactment of Section 409A as part of the American Jobs Creation Act of 2004 (the "2004 Tax Act").[11] Section 409A does not address any of the concerns in answers "a"

---

5. *See infra* text accompanying note 45 (providing a similar situation in Example #1).

6. *See infra* text accompanying notes 17-18 (describing the issue of grossly excessive executive compensation).

7. *See infra* notes 20-36 and accompanying text (describing the underlying rationales for the absence of arm's length bargaining for executive compensation).

8. *See infra* notes 37-39 and accompanying text (describing how boards and executives do not adequately disclose top executives compensation).

9. *See infra* notes 40-42 and accompanying text (describing the flaws of marginal tax rates).

10. *See infra* text accompanying note 43 (describing how Section 409A affects only NQDC).

11. American Jobs Creation Act of 2004, Pub. L. No. 108-357, § 885(a), 118 Stat. 1418

through "d" listed above. Instead, Section 409A only addresses answer "e" and sets a series of bizarre traps for an unwary employee who will be compensated in the future for services already rendered. Section 409A generally applies regardless of (i) the size of the entity paying the compensation; (ii) whether the entity paying the compensation is public or private;[12] (iii) the total amount of compensation earned by the employee; (iv) the amount of compensation deferred;[13] or (v) the employee's job classification. Section 409A will have a relatively minor impact on well-advised, highly compensated executives at publicly-traded companies, but will be a trap for uninformed, smaller employers and their employees.

At its peak, Enron ranked seventh among the Fortune 500 and had over 25,000 employees worldwide.[14] Enron's collapse focused the public eye on Corporate America and may be credited as the turning point for the regulation of accounting practices and corporate governance. In response to widespread publicity regarding perceived tax and compensation abuses at Enron, the Congressional Joint Committee on Taxation (the "Joint Tax Committee") conducted multiple hearings and issued a 732-page report (the "Enron Compensation Report").[15] The Enron Compensation Report published information about the actual compensation practices of a major public corporation, including information that would normally be unavailable to the general public.[16]

After the collapse of Enron, several other major corporate scandals hit the newsstands. The publicity surrounding Enron, WorldCom, Global Crossing,

---

(2004).

12. The rules of Section 409A apply equally to public and private companies, except that one provision requires that payments following a "separation from service" cannot begin for at least six months in the case of a "key employee . . . of a corporation any stock in which is publicly traded on an established securities market or otherwise." *See* I.R.C. § 409A(a)(2)(B)(I) (2005).

13. An employer may distribute an employee's entire NQDC account balance in a lump sum upon her separation from service if the entire amount is under $10,000. *See* I.R.C. § 409A(a)(3); H.R. REP. NO. 108-755, at 731-32 (2004) (Conf. Rep.) [hereinafter Conference Agreement]; Prop. Treas. Reg. § 1.409A-3(h)(2)(iv)(A)(3), 70 Fed. Reg. 57930, 57980 (Oct. 4, 2005); I.R.S. Notice 2005-1, 2005-2 I.R.B. 274, 283.

14. Enron Compensation Report, *supra* note 2, at 5.

15. The Joint Tax Committee staff investigation was conducted in response to a letter from Senators Max Baucus and Charles E. Grassley of the Senate Committee on Finance. Enron Compensation Report, *supra* note 2, at 1.

16. The Enron Compensation Report states:

> [T]he Joint Committee staff believes that its investigation provides valuable analysis of Enron's structured transactions and compensation structures and provides important recommendations and findings for improvements to the Federal tax system. More generally, the Joint Committee staff believes the Report provides significant insights into a corporation's tax and compensation activities that typically are unavailable to those outside the company.

Enron Compensation Report, *supra* note 2, at 3.

Tyco, HealthSouth, and others created an environment conducive to fundamental reforms in areas such as corporate governance, accounting, disclosure, compensation, and tax practice.

In the area of executive compensation, the three topics in answers "a" through "c" above were ripe for reform. First, "[b]etween 1992 and 2000, the average real (inflation-adjusted) pay of chief executive officers ("CEOs") of S&P 500 firms more than quadrupled, climbing from $3.5 million to $14.7 million."[17] In February 2004, Warren Buffett stated, "In judging whether Corporate America is serious about reforming itself, CEO pay remains the acid test. To date, the results aren't encouraging."[18]

A frequent response to claims of excessive CEO compensation is that the compensation is established by arm's-length bargaining, and companies are paying the minimum amount necessary to retain the employee's services. For example, while many argued that baseball player Alex Rodriguez' compensation of $252 million over 10 years was excessive,[19] one could forcefully argue that the team's owners had every incentive to bargain aggressively with Rodriguez, and that the compensation paid to Rodriguez represents the fair market value of his services. While the "arm's-length bargaining" argument may justify compensation paid to professional athletes and executives of closely-held businesses,[20] this argument, as described in the following paragraphs, falls apart in the case of top executives at publicly-traded corporations.[21]

---

17. LUCIAN BEBCHUK & JESSE FRIED, PAY WITHOUT PERFORMANCE: THE UNFULFILLED PROMISE OF EXECUTIVE COMPENSATION 1 (2004).

18. Letter from Warren Buffett, Chief Executive Office, Berkshire Hathaway Inc., to shareholders of Berkshire Hathaway Inc. (Feb. 2004), *reprinted in* BEBCHUK & FRIED, *supra* note 17, at xiii.

19. *See* Ian O'Connor, *Money for Nothing*, THE JOURNAL NEWS (Westchester County, NY), Nov. 18, 2004, at 1C (noting that the Yankees assumed $112 million of Rodriguez' $252 million contract with the Texas Rangers, while also agreeing to pay $189 million to Derek Jeter, $120 million to Jason Giambi, and $88.5 million to Mike Mussina); *see also* Howard Bryant, *Glory Eludes A-Rod*, BOSTON HERALD, May 30, 2005, at 92 (discussing Alex Rodriguez' $252 million contract).

20. If the executive has a family relationship with the owners of a closely-held business, or has substantial ownership in the corporation, then the "arm's-length bargaining" argument is not persuasive. *See* Barbourville Brick Co. v. Comm'r, 37 T.C. 7, 11-13 (1961) (concluding that amounts the corporation paid to the majority shareholder were non-deductible dividends and were not reasonable compensation deductible under section 162 of the I.R.C.).

21. A potential response to claims that top executives at publicly-traded corporations are overpaid, thus injuring the shareholders, is that people voluntarily become shareholders, that shareholders as a group are economically sophisticated, and that they can protect themselves. But with the proliferation of 401(k) plans and similar arrangements, over 60% of all U.S. households directly or indirectly invest in the stock market. *See* Raksha Arora, *Portrait of the American Investor*, THE GALLUP POLL TUESDAY BRIEFING, May 4, 2004, at 55 ("Given that there are approximately 105 million households in the United States . . . poll numbers suggest that more than 64 million American households are invested in the stock market, either through

Second, boards of directors of publicly-traded companies have many reasons for not bargaining at arm's-length with top executives when establishing executive compensation. "Directors have had various economic incentives to support, or at least go along with, arrangements favorable to the company's top executives."[22] Incentives may include the top executive's ability: (i) to channel business now or in the future to the directors and their related entities; (ii) to impact the slate of persons nominated to the Board of Directors; and (iii) to influence the amount of fees and other perks paid by the corporation to the directors. Further,

> Various social and psychological factors—collegiality, team spirit, a natural desire to avoid conflict within the board team, and sometimes friendship and loyalty—have . . . pulled board members in that direction. Although many directors own shares in their firms, their financial incentives to avoid arrangements favorable to executives have been too weak to induce them to take the personally costly, or at the very least unpleasant, route of haggling

---

individual stock, a mutual fund, or a self-directed 401(k) or IRA."); Virginia Baldwin Gilbert, *Investor's Roller Coaster Ride Taught Him to Diversify; The Bear Market Advisers Agree that Long-Term Investing is the Safest Way to Go*, ST. LOUIS POST-DISPATCH, Mar. 18, 2001, at A14 ("The number of 401(k) participants alone has grown 230 percent since 1987—to 43 million from 13 million . . . . [P]articipants in all defined-contribution plans in the United States number close to 60 million, investing $2.5 trillion."). Also, if the Social Security System is revised to create private investment accounts for all workers, presumably almost all employees will become shareholders of public companies. *See* Ellyn Ferguson, *Smith Keeps Options Open on Social Security*, GANNETT NEWS SERVICE, July 7, 2005, at ARC ("[President] Bush has called for putting a portion of each worker's Social Security taxes into private accounts that would be invested in the . . . stock market.").

Further, existing structures discourage boards of directors from bargaining at arm's-length with executives, *see infra* Part VI.A.1., and shareholders lack effective mechanisms for addressing the problem of excessive compensation payments to top executives. *See* BEBCHUK & FRIED, *supra* note 17, at 45 ("[S]hareholders have sought to constrain executive compensation arrangements in three ways—by (1) suing the board, (2) voting against employee stock option plans, and (3) putting forward shareholder resolutions. . . . [N]one of these methods has imposed significant constraints on executive pay."). For example, Walt Disney Company shareholders were unsuccessful in challenging an executive compensation scheme, even though the Delaware Chancery Court Judge stated, "[Michael] Eisner stacked his (and I intentionally write 'his' as opposed to 'the Company's') board of directors with friends and other acquaintances who, though not necessarily beholden to him in a legal sense, were certainly more willing to accede to his wishes and support him unconditionally than independent directors." David Lieberman, *Disney Wins Ruling in Lawsuit on Ovitz Pay*, USA TODAY, Aug. 10, 2005, at 01B. The court, in rejecting shareholder claims, "ruled that company directors adequately looked out for shareholders' interests in 1995 when they endorsed CEO Michael Eisner's decision to hire his former pal Michael Ovitz as president—then, 14 months later, approved his firing with a $130 million severance package." *Id.* In fact, the Chairman of the SEC has expressed frustration over how the problem can be addressed by either shareholders or the government. BEBCHUK & FRIED, *supra* note 17, at 189.

22. BEBCHUK & FRIED, *supra* note 17, at 4.

with their CEOs. Finally, limitations on time and resources have made it difficult for even well-intentioned directors to do their pay-setting job properly.[23]

Often, the directors' primary restraint when negotiating executive compensation is merely the potential outrage of the shareholders and the general public if the arrangement is extraordinary.[24] For example, when Citigroup's CEO announced that he was contemplating retiring prior to the termination of his employment agreement and starting a competing firm, the board of directors declined to grant him free use of Citigroup's jet fleet for life because "[i]t wouldn't look good."[25] In several situations, as charted below, boards of directors have apparently miscalculated and authorized compensation arrangements that have been considered outrageous, at least by the media.

Public Corporation	**CEO Compensation**
Tyco International Ltd.	L. Dennis Kozlowski allegedly received $430 million for "inflating the value of Tyco stock," another $170 million in disputed bonuses, the use of an $18 million Manhattan apartment, a $2 million birthday party on the island of Sardinia, $30,000 opera glasses, a $16,000 dog-shaped umbrella stand, and

23. *Id.*

24. *See id.* at 5 (stating that the "outrage" of the shareholders or the general public can result in embarrassment or damage to reputation for the directors).

25. Mitchell Pacelle & Monica Langley, *Weill's Perks Create Obstacles; Wary of Negative Publicity, Citigroup Board Balks at Paying for Corporate Jet, Security Detail*, WALL ST. J., July 20, 2005, at C1.

26. A.J. Carter, *Lessons of the Boys of October*, NEWSDAY, Oct. 3, 2003, at A46 ("Kozlowski went on trial for looting the company treasury for amenities . . . such as the umbrella stand shaped like a dog ($16,000) and the party in Sardinia for which Jimmy Buffet [sic] was flown in to sing "Happy Birthday" to the guest of honor, Kozlowski's wife ($2 million)."); Mike Drummond, *Firms with Solid Ethical Footing can Still Recover when Leaders go Astray*, THE CHARLOTTE OBSERVER, May 2, 2005 ("[Tyco] was brought low a few years ago with tales of extravagant and ribald parties featuring X-rated ice sculptures, Kozlowski's purchase of a $6,000 shower curtain, and other excesses."); Samuel Maull, INTELLIGENCER JOURNAL (Lancaster, PA), Mar. 17, 2004, at 8 ("Kozlowski . . . and [the top financial officer, Mark H. Swartz] are accused of stealing $170 million by taking unauthorized bonuses and abusing company loan programs, and netting an additional $430 million by manipulating Tyco stock prices.") (discussing the $18 million Manhattan apartment, the $30,000 opera glasses, and the $6,000 shower curtain in the maid's bathroom). Kozlowski also stated that he "didn't notice that $25 million in 1999 compensation was not reflected on his W-2." Jim Wooten, *Thinking*

	a $6,000 shower curtain in the maid's bathroom.[26]
Morgan Stanley	Steve Crawford, age 41, received $32 million as a termination payment after serving as co-president of the company for three months.[27]
Morgan Stanley	Philip Purcell received $113 million in stock and cash for leaving the company.[28]
General Electric	Jack Welch received "$120 million in his last year as CEO alone."[29]

Right: Court Rulings, Winn-Dixie, Panhandlers, ATLANTA JOURNAL-CONSTITUTION, June 24, 2005, at A17; *see also* Editorial, HONOLULU ADVERTISER, Mar. 21, 2004, at 2 ("[F]ormer Tyco CEO Dennis Kozlowski made nearly $467 million in salary, bonuses and stock during his four years running the company into the ground."). Further, Tyco shareholders lost $80 billion during Kozlowski's reign. *See* Ellen Frank, *The Great Stock Illusion*, DOLLARS & SENSE, Nov. 1, 2002, at 14 ("Kozlowski defended this windfall [his compensation] with the claim that his leadership had 'created $37 billion in shareholder wealth.' By the time Kozlowski quit Tyco under indictment for sales tax fraud in 2002, $80 billion of Tyco's shareholder wealth had evaporated.").

27. Greg Farrell, *Morgan Stanley sued over payments to former execs*, USA TODAY, July 20, 2005, at 6B ("Crawford . . . quit last week to take advantage of a big payout that would [have] vanish[ed] if he stayed beyond August." Further, "Jack Coffee, an expert in securities law at Columbia University . . . says . . . the $32 million windfall awarded to Crawford after only three months as a senior executive 'looks disingenuous'."); Charles Gasparino with Nicole L. Joseph, *Good News: You're Fired*, NEWSWEEK, July 25, 2005, at 48 ("Many big Morgan shareholders—who've watched the firm's stock fall from more than $100 a share to about $54 in recent years—are *outraged*, particularly about the pay for Crawford, who was never a star banker or big moneymaker at the firm." (emphasis added)); *see also* Ann Davis, *Changing of Guard At Morgan Stanley Is Moving Forward*, WALL ST. J., July 15, 2005, at C1 ("Mr. Crawford's departure . . . triggered a widely criticized $32 million exit package . . . .").

28. Farrell, *supra* note 27, at 6B; *see also* Gasparino, *supra* note 27, at 48 (indicating that $44 million of the $113 million was paid in cash); Pacelle, *supra* note 25, at C4 (stating that Morgan Stanley awarded $44 million of exit pay, $62.3 million in stock and option awards, and $250,000 of NQDC per year for life); John Dimsdale, *Marketplace Morning Report: Pension fund suing Morgan Stanley's board of directors for awarding excessive pay packages to resigning managers*, (Minn. Public Radio broadcast, July 20, 2005) ("The shareholder group's lawyer . . . calls those severance packages '*wretched excess*.'" (emphasis added)).

29. Michael B. Dorff, *Does One Hand Wash the Other? Testing the Managerial Power and Optimal Contracting Theories of Executive Compensation*, 30 J. CORP. L. 255, 256 (2005); Pacelle, *supra* note 25, at C1 (stating that Mr. Welch became a "symbol of corporate excess"); John L. Sweeney, *The Foxes Are Still Guarding the Henhouse*, LOS ANGELES TIMES, Sept. 19, 2003, at B13 ("[M]any of today's executive compensation packages are excessive. The lavish retirement lifestyle of . . . Jack Welch—including corporate jets, posh housing, country club memberships and wine and laundry services . . . is but one example.").

Walt Disney Co.	"Michael Eisner took home nearly $600 million in stock option profits from Walt Disney Co. in a single year . . . ."[30]
Hewlett Packard	Carly Fiorina received "six years of excessive salary, bonuses and stock options, . . . $110 million-plus in 'special' one-time bonuses,"[31] and a $21 million severance package.[32] HP awarded Fiorina a "$567,000 discretionary bonus in mid-December" 2004, just before her dismissal in February 2005.[33] During Fiorina's reign, HP's market equity dropped nearly 60%.[34]
May Company	Gene Kahn received "$9 million [as a] 'long goodbye' gift . . . after [May Co.] profits dropped 50 percent in four years[.] Even more amazing was giving Mr. Kahn two years of salary, totaling $3 million, in return for a non-compete agreement. With performance like his, why would you care?"[35]

Clearly a disconnect between directors and shareholders exists. For example, when several shareholders at the Hewlett Packard annual shareholder meeting complained about the CEO's compensation, and one shareholder asked, "Why do CEOs need this kind of money?" the Chairman of the Board simply replied, "The board feels it is not in the best interest to have protracted

30. Susan Chandler, *'Golden Hellos' Still Glitter*, CHI. TRIB., July 19, 2005, at C1.

31. William Brown, *Fiorina's Laughing with her Millions*, SAN JOSE MERCURY NEWS, Feb. 12, 2005, at B1 ("Lest we become too euphoric about the departure of Carly Fiorina from Hewlett-Packard, let me remind you that the damage has all been done, and Carly is laughing all the way to the bank.").

32. Therese Poletti, *Owners Get Few Details on HP's CEO Search*, SAN JOSE MERCURY NEWS, Mar. 17, 2005, at BU1.

33. Joann S. Lublin, *CEO Bonuses Rose 46.4% at 100 Big Firms in 2004*, WALL ST. J., Feb. 25, 2005, at A1 ("'You have to wonder how directors justify paying big bonuses and a few months later firing a CEO,' says Carol Bowie, director of governance-research services at the Investor Responsibility Research Center in Washington.").

34. H.D. Maynard, Editorial, *Failure Shouldn't Pay*, ST. LOUIS POST-DISPATCH, July 22, 2005, at B8.

35. *Id.*

negotiations with people when they leave the company[.]"[36] Would the Board negotiate differently if they were spending their own money?

Third, reporting executive compensation to shareholders, potential investors, and the general public under the current rules can be complex, confusing, and incomplete.[37] In discussing executive compensation, the director of the SEC's Corporation-Finance Division stated that "[t]oo many issuers and their advisers have followed a *pattern of opaque or unhelpful disclosure*."[38] Because boards of directors lack appropriate incentives to bargain at arm's-length with top executives, and instead merely seek to avoid shareholder and public outrage, boards and executives have a common objective to camouflage executive compensation.[39] Thus, boards and executives may seek out methods of compensation that do not require full disclosure, even though such methods of compensation fail to provide appropriate incentives to executives, or are otherwise economically inefficient.

Regarding the topic in answer "d" above,[40] both before and after the Enron scandal, the general lack of progressivity of income tax rates applicable to all levels of income (including executive compensation) runs counter to the fundamental tax principle of taxing on the "ability to pay."[41] A lower- or middle-management employee earning $185,000 per year is taxed at almost the same marginal income tax rate as the CEO of a public company who is paid

---

36. Poletti, *supra* note 32.

37. One commentator recently stated:

> What are the current problems with executive-comp disclosure? For one, elements of the pay package are spread across numerous SEC filings. The main information is in the annual proxy, filed generally sometime in the spring. But other essentials can be found in other filings: in interim filings, if a payment contract is amended, or in the forms insiders must file when they buy or sell shares or exercise options. Often, the proxy is already out of date.
>
> . . . .
>
> Another problem is that there is no one table that includes all the compensation, including salary, bonus, equity grants, accumulated deferred compensation, pension plans, perks and tax benefits and anything else a company sees fit to lavish on its top dogs. And one helpful line is nowhere to be found: "Total Compensation."

Jesse Eisinger, *Follow the CEO's Money*, WALL ST. J., Feb. 16, 2005, at C1. As an example of the complexity, the Wall Street Journal hired an actuary and the Boston Globe hired an executive-pay consultant to calculate the compensation paid to Gillette CEO James Kilts in connection with Procter & Gamble's purchase of Gillette—one calculated $185 million; the other calculated $173 million. *Id.*

38. *Id.* (emphasis added).

39. BEBCHUK & FRIED, *supra* note 17, at 5 ("The critical role of outsiders' perception of executives' compensation and the significance of outrage costs explain the importance of yet another component of the managerial power approach: 'camouflage.'").

40. *See supra* note 9 and accompanying text.

41. JOEL NEWMAN, FEDERAL INCOME TAXATION: CASES, PROBLEMS, AND MATERIALS 15 (2d ed. 2002) ("Fairness in taxation is often discussed in terms of ability to pay. Those who are able to pay more taxes should pay more; those who are able to pay less taxes should pay less.").

$10 million per year, even though the $10 million executive can pay tax at a much higher marginal rate.[42] While some lower- and middle-management employees may derive an odd sense of pride from being taxed at roughly the same marginal income tax rate as Kenneth Lay, Bernie Ebbers, Dick Grasso, and Dennis Kozlowski, most would view this imbalance as a fundamental flaw of the U.S. income tax system.

Instead of tackling these fundamental issues and adopting finely-tuned rules to address actual abuses at publicly-traded companies, Section 409A disappoints in three major ways: (1) Section 409A only addresses answer "e" above, which merely concerns the *timing* of a small portion of a typical executive's total compensation package (NQDC represented less than 5% of the total compensation provided to the top 200 executives at Enron);[43] (2) Section 409A was designed to stop an abuse that did not exist; and (3) instead of being finely-tuned, Section 409A applies to all employers equally, making it a minor annoyance for well-advised publicly-traded companies and their highly-paid executives, but a dangerous trap for unsophisticated employers and their employees.

First, Section 409A merely establishes new timing rules for the payment of NQDC arrangements. In such an arrangement, a service provider (e.g., an executive) and the service recipient (e.g., a corporation) agree that the service provider will receive a portion of her compensation in a future tax year for services rendered currently. Section 409A prohibits the parties from *changing the timing* of payments to the service provider after the services have been rendered and the parties have agreed when the deferred compensation will be paid. Basically, Section 409A prohibits the service provider and the service recipient from agreeing to accelerate the payment of the NQDC after they have previously agreed on a payment schedule.[44]

> *EXAMPLE #1.* Helpful Helen agreed with her employer at the beginning of 2000 that, as a part of her savings plan for retirement, $20,000 of her compensation earned during 2000 would be deferred until her death, disability, or termination of employment. In 2006, the family finances are not going as well as expected, and Helpful Helen would like to receive that deferred amount promptly to help pay for part of her children's tuition. Prior to the enactment of Section 409A, Helpful Helen and her employer could agree to distribute the money to Helen early.[45]
>
> But after the enactment of Section 409A, Helen will incur tremendous tax costs if she receives the money before her death, disability, or termination

---

42. In 2005, a married couple filing jointly hits the 33% marginal income tax bracket with taxable income of $182,800, and hits the highest tax bracket of 35% with taxable income of $326,450. *See* I.R.C. § 1(i) (2005); RIA Federal Tax Handbook 2005, ¶1103, page 51 (rates for 2005).

43. *See infra* note 94.

44. I.R.C. § 409A (a)(3).

45. *See generally* Metcalfe v. Comm'r, 43 T.C.M. (CCH) 1393 (1982) (holding that an employee's rights to deferred amounts of compensation could not be forfeited).

of employment.[46] Section 409A prevents Helpful Helen from using her NQDC to help pay tuition, even if her employer is willing to accelerate the payment. As this example illustrates, Section 409A establishes mere timing rules and fails to address fundamental issues involved in overcompensating executives.

Second, the primary rules of Section 409A appear to originate from the Enron Compensation Report,[47] which provides facts that demonstrate the fundamental problems of executive compensation at publicly-traded companies (discussed in answers "a" and "b" of the multiple-choice question at the beginning of this Article).[48] But rather than address these fundamental problems, the Section 409A rules focus on an abuse that did not exist in the Enron debacle.

The Enron Compensation Report focuses on Enron's distributions of NQDC funds to top executives prior to Enron's bankruptcy and asserts that, as a result, top executives obtained and retained their NQDC, while lower-level employees lost almost all their retirement savings in the Enron collapse.[49] In response, the Enron Compensation Report recommends eliminating all flexibility over the timing of payments once compensation has been deferred,[50] and Section 409A enacts that recommendation. But because bankruptcy laws enacted long before the Enron bankruptcy and subsequent enactment of Section 409A allow a bankruptcy trustee to recover NQDC payments made to a top executive within one year of the company's bankruptcy,[51] the perceived abuse that triggered the enactment of Section 409A did not exist.

Third, Section 409A applies to all employers, rather than tailoring provisions to target particular employers who are more likely to abuse situations. This deficiency does not result from Congress' inability to pinpoint abusive situations; in previous years, Congress has demonstrated that it can fine-tune its tax rules to prevent compensation abuses. In 1993, Congress enacted the "$1 Million Cap," which was designed to limit the compensation of certain top executives.[52] Specifically, the $1 Million Cap[53] applies to the CEO

---

46. *See infra* note 229 and accompanying text (providing Example #4, which demonstrates the harshness of a violation of Section 409A).

47. *See infra* notes 59-64 and accompanying text.

48. *See infra* Part II.C (discussing the grossly excessive pay of Enron's 200 highest paid employees and the failure of Enron's board to negotiate executive compensation at arm's-length).

49. Enron Compensation Report, *supra* note 2, at 37 ("[M]any executives . . . reaped substantial gains from their compensation arrangements. Enron's rank and file employees in many cases lost virtually all of their retirement savings because they believed statements made by Enron's top executives up to the very end that Enron was viable and that Enron's stock price would turn around.").

50. *See infra* notes 146-48 and accompanying text.

51. *See infra* Part II.F.1.

52. *See* I.R.C. § 162(m) (2005).

53. *See id.* § 162(m)(2).

and the four other highest paid executives at each publicly-traded company.[54] Similarly, the "golden parachute" rules adopted in 1984[55] only apply[56] to officers, shareholders, or highly-compensated officers[57] at publicly-traded companies who receive more than three times their prior average annual compensation when they terminate employment with the corporation in connection with a change of control.[58]

Although perceived abuses at Enron, including Ken Lay's deferral of $32 million in one year and the deferral of millions of dollars by other top Enron executives,[59] inspired Congress to enact Section 409A, the section generally applies to *all* employers and employees,[60] regardless of the amounts of money involved.[61] In sharp contrast to the $1 Million Cap[62] and the golden parachute rules,[63] which were finely-tuned to apply when the risk of abuse was high,

---

54. *See id.* § 162(m)(3). *See generally* Gerald A. Kafka, *Reasonable Compensation, in* 390-3d TAX MANAGEMENT PORTFOLIOS, at A-37 to -40 (BNA 2005) (discussing the $1 million limit on deductible compensation and compensation that is not subject to this limit).

55. *See* I.R.C. § 280G.

56. *See id.* § 280G(b)(5)(A)(ii). A corporation is not allowed a tax deduction for an "excess parachute payment." *Id.* § 280G(a). An excise tax of 20% is imposed on an executive who receives an excess "parachute payment." *Id.* § 4999(a). This term, however, does not include payments by a "small business corporation," *id.* § 280G(b)(5)(A)(i), or payments by a corporation if, immediately before a change of control, "no stock in such corporation was readily tradeable on an established securities market or otherwise, and the shareholder approval requirements . . . are met . . . ." *Id.* § 280G(b)(5)(A)(ii).

57. *Id.* § 280G(c)(2). For these purposes, "the term 'highly-compensated individual' only includes an individual who is . . . a member of the group consisting of the highest paid 1 percent of the employees of the corporation or, if less, the highest paid 250 employees of the corporation." *Id.*

58. *Id.* § 280G(b)(2)(A)(ii).

59. *See infra* notes 93-97 and accompanying text; *see also* Stephen W. Skonieczny & Abigail B. Pancoast, *Rules Governing Executive Compensation in Rare State of Flux*, N.Y. L.J., Feb. 23, 2004, *available at* http://www.dechert.com/library/execcomprules.pdf ("Congress has recently introduced several proposals to change the tax treatment of deferred compensation arrangements. . . . [T]hese proposed changes stem in part from the recent corporate scandals and a related public perception of excessive executive pay.").

60. There are, however, a few provisions of Section 409A that are finely-tuned. For example, when a payment is made upon "separation from service" in the case of a key employee of a publicly-traded company, no NQDC payment can be made within six months of the separation from service without violating Section 409A. I.R.C. § 409A(a)(2)(B)(i). The Proposed Treasury Regulations provide that the entire amount under a NQDC plan can be distributed immediately upon a separation from service if the amount is under $10,000. *See* Prop. Treas. Reg. § 1.409A-3(h)(2)(iv)(A)(3), 70 Fed. Reg. 57930, 57980 (Oct. 4, 2005); I.R.S. Notice 2005-1, 2005-2 I.R.B. 283.

61. *But see supra* notes 13 and 60 (discussing an exception that permits an entire account balance under $10,000 to be paid to the employee immediately upon a separation from service).

62. *See infra* notes 446-51 and accompanying text (discussing the $1 Million Cap and Section 162(m)).

63. *See supra* notes 55-58 and accompanying text (discussing the "golden parachute"

Section 409A applies to mom-and-pop restaurants and charitable organizations, as well as Fortune 500 companies.[64]

While Section 409A will make the payment provisions of NQDC plans less flexible, highly-paid executives will likely be content to wait for the deferred portion of their compensation because, in many cases, they do not need the money anyway. Section 409A imposes no limit on the amount of compensation that one can defer, and imposes no additional procedural or administrative requirements that would encourage directors to bargain at arm's-length over executive compensation. Well-informed, sophisticated employers, including publicly-traded companies, will be able to design or redesign NQDC plans to comply with Section 409A. Thus, Section 409A likely will be a minor annoyance for sophisticated employers and their executives. But Section 409A and the related Internal Revenue Service ("IRS") guidance[65] will present the uninformed with a fiendishly complex web of detailed rules that will trap many small employers and their employees. Preliminary reports indicate that the initial IRS regulations interpreting Section 409A will exceed 100 pages.

This Article considers executive compensation, NQDC, and Section 409A in five sections. Section II of this Article describes the role of NQDC in the Enron scandal and subsequent corporate compensation scandals, and describes a fundamental flaw of the Enron Compensation Report. Specifically, the primary NQDC "abuse" identified in the Enron Compensation Report and targeted through enactment of Section 409A did not exist. Instead, Section 409A did not address the real abuses that are revealed in the Report, including the fundamental issues raised at the beginning of this Article. Section III of this Article briefly describes NQDC arrangements in general and their typical function as a natural arrangement driven by economic necessity rather than tax considerations. This section also explores the basic tax rules that governed these arrangements before enactment of Section 409A. Section IV of this Article discusses the penalties for violating Section 409A, the types of arrangements subject to Section 409A, the major changes created by Section 409A, and the impact that these new rules will have on employees. Section V

---

rules and section 280G of the I.R.C.).

64. A 1996 New York Times article discussing the prevalence of NQDC plans stated, "From Ollie's Station Restaurant in Tulsa, Okla., to Bubba's Produce in New Orleans, thousands of small-business owners are creating plans to defer compensation...." Christopher Drew & David Cay Johnston, *Special Tax Breaks Enrich Savings of Many in the Ranks of Management*, N.Y. TIMES, Oct. 13, 1996, at A1. The charitable organizations cited as using NQDC plans include: the Knights of Columbus in New Haven, Connecticut; Mount Sinai Hospital in New York City; the National Resource Defense Council in New York City and Washington (an environmental protection organization); and the Boy Scouts of America Greater New York Councils. *Id.*

65. Because of the tremendous complexity of Section 409A and the almost immediate effective date upon enactment, the IRS announced that it would need to issue a "*series* of guidance [documents]" within the first year after Section 409A's enactment. I.R.S. Notice 2005-1, 2005-2 I.R.B. 274 (emphasis added).

of this Article identifies particularly inappropriate aspects of Section 409A that should be changed or eliminated retroactively. Section VI of this Article discusses various policy themes that should be considered in future attempts at compensation reform.

II. THE ROLE OF NQDC IN VARIOUS CORPORATE SCANDALS, THE ENRON COMPENSATION REPORT, AND ITS FUNDAMENTAL FLAW

A. The Role of NQDC: Just Another Part of the Compensation Problem

NQDC is merely compensation paid in a year after it is earned, and in many cases it is just one piece of the compensation package designed for executives.[66] For example, from 1998 to 2000, NQDC represented only 5% of the total compensation of Enron's top 200 executives.[67] The fundamental problems of executive compensation (the "Fundamental Compensation Problems") include:

- Grossly excessive amounts of compensation;
- Lack of incentives for directors at public companies to bargain at arm's-length with executives over compensation;
- Inadequate disclosure to investors, creditors, and others; and
- Taxing "small fish" at the same marginal income tax rates as "fat cats."

Thus, whenever these problems are addressed, NQDC should be included as part of the reform. For example, if Congress adopts procedural rules to encourage directors at public companies to carefully scrutinize executive compensation arrangements, those rules should apply to NQDC arrangements, as well as other compensation arrangements. But because NQDC is merely one piece of the compensation puzzle, any attempt to cure the Fundamental Compensation Problems by addressing only NQDC is akin to one spending all of her time organizing a closet when she needs to clean the entire house.[68]

---

66. The compensation package for a top executive typically would include fixed salary, bonuses, stock options, severance pay, qualified plan benefits, life insurance, and fringe benefits.

67. *See infra* note 94.

68. There are certain ways in which NQDC is treated differently than compensation in general. For example, NQDC that is not payable until after termination of employment does not count in calculating the $1 million compensation cap of Section 162(m). Because Section 162(m) only allows a corporation to deduct $1 million of fixed compensation to each of its five highest paid executives, "companies encouraged executives to postpone taking the amount of their pay in excess of $1 million until after they left the company or retired." Ellen E. Schultz & Theo Francis, *Buried Treasure: Well-Hidden Perk Means Big Money for Top Executives*, WALL ST. J., Oct. 11, 2002, at A1; *see also* BEBCHUK & FRIED, *supra* note 17, at 102 ("Some firms require that managers receiving salary in excess of $1 million, which would otherwise be

B. NQDC in Recent Corporate Scandals

Recent media coverage of outrageous executive compensation reveals that the scandals frequently involve use and abuse of deferred compensation plans. Perhaps the kingpin of NQDC scandals is Richard Grasso, former Chairman of the New York Stock Exchange, a nonprofit organization.[69] While Mr. Grasso's total compensation was enormous—from 1995 to 2003, he received about $193 million in salary and pension payments[70]—the firestorm occurred following the public disclosure of the additional $139.5 million payable immediately to Mr. Grasso in connection with his ouster in September 2003, and the additional $48 million to be paid in future years. Grasso argued that the $48 million "did not need to be disclosed at the time [of his ouster] because it was comprised of payments in the future."[71] Apparently Mr. Grasso is in good company, as revealed through media coverage of several other executive compensation scandals:

- Kenneth Lay deferred $32 million of compensation in one year.[72] His total compensation during Enron's last three years was approximately $200 million.[73]

---

nondeductible under Section 162(m)... defer the excess."). In addition, only the portion of the interest earned on NQDC in excess of a market rate, as determined by the government, is disclosed for securities purposes. *See id.* at 106 ("The SEC requires firms to include in the compensation table for each executive the above-market interest earned that year on deferred compensation. In the case of a guaranteed interest rate, 'above-market' interest is defined as returns in excess of 120 percent of the applicable federal rate (AFR) used by the [IRS] at the time the guaranteed interest rate is set...."); Schultz & Francis, *supra*, at A1 ("Companies are required to report only guaranteed above-market interest paid annually into the accounts of the five highest-paid executives. They are not required to disclose interest paid below that rate or any gains pegged to stock funds, hedge funds or other investments with fluctuating returns."). Also, arguably the ability of the employee to "defer" income taxation on NQDC can create a "timing" benefit that might be addressed by taxing the employee on an imputed interest income amount. *See* Daniel I. Halperin, *A Fairer and More Effective Approach to Deferred Compensation*, 107 TAX NOTES 1187, 1189 (2004) ("[D]eferral can be freely allowed as long as we impose a special tax on investment income at the rate appropriate to the employee." (citing Daniel I. Halperin, *Interest in Disguise: Taxing the Time Value of Money*, 95 YALE L.J. 506, 539-50 (1986))). While "tweaking" these specific issues involving NQDC may be appropriate, the minor problems that are unique to NQDC should not delay the important work of addressing the Fundamental Compensation Problems.

69. Mr. Grasso has been called "the poster child of excessive pay." Patrice Hill, *Public Skeptical of Corporate Corruption Crackdown*, THE WASHINGTON TIMES, Apr. 19, 2004, at A01.

70. Aaron Lucchetti & Susanne Craig, *NYSE Report Shows Excesses Under Grasso*, WALL ST. J., Feb. 3, 2005, at C1 (stating that compensation experts report that approximately $100 million of that compensation was "excessive."). In addition, Mr. Grasso's secretary was paid $240,000 a year, and his two drivers were each paid $130,000 a year. *Id.*

71. *Id.* A New York State Supreme Court judge ruled that Mr. Grasso must return up to $100 million of his compensation. Landon Thomas, Jr., *Ex-Stock Exchange Chief Told to Return Up to $100 Million*, N.Y. TIMES, Oct. 20, 2006, at A1.

72. Enron Compensation Report, *supra* note 2, at 604 n.1817.

- Jack Welch became an icon for corporate greed as a result of information revealed during his divorce.[74] He had a NQDC account balance of approximately $21.1 million at General Electric in 2001.[75] The deferrals and the interest on his NQDC account balance totaled approximately $2.87 million, or 17.8% of his 2001 compensation.[76]
- In 2001, Tyco contributed "principal" of $397,450, and "above-market interest"[77] of $219,543 to Dennis Kozlowski's NQDC plan.[78]
- Fannie Mae paid its former CEO, Franklin Raines, $20 million in salary, bonus, and stock in the year before he was fired.[79] Further, Fannie Mae awarded him $26 million in severance pay plus a NQDC benefit worth almost $1.4 million per year even though Fannie Mae lost almost $9 billion under his leadership.[80]
- A study conducted on the salaries of non-profit CEOs revealed, "The United Way of America has changed its pension plan to bar lump-sum payouts, after reporting that it had made a $1.5-million pension payment to its former chief executive, Betty Stanley Beene, when she departed after four years in the job."[81]
- Morgan Stanley's board of directors was widely criticized for its compensation payments to CEO Philip Purcell upon his ouster. Purcell received $44 million in cash as severance, $62.3 million in stocks and options, and a comparatively small $250,000 per year for life as NQDC.[82]

---

73. Barrie McKenna, *The House that Enron Built*, THE GLOBE AND MAIL (Toronto, Ontario), Feb. 2, 2002, at F4 (stating that former Enron chairman Kenneth Lay "[h]elped to found Enron in 1985 and since 1999 has received about $200-million . . . in salary, stock and other compensation").

74. Pacelle, *supra* note 25, at C1 (stating that Mr. Welch became a "symbol of corporate excess").

75. Schultz & Francis, *supra* note 68, at A1 (stating that Welch retired at age 66 in September 2001; during that year he received $3.4 million in salary, a $12.7 million bonus, $2.53 million of "interest" on his NQDC account balance, and GE contributed an extra $340,375 to Mr. Welch's NQDC plan).

76. *Id.*

77. This includes only interest in excess of 5.9%.

78. Schultz & Francis, *supra* note 68, at A1. If Tyco was crediting Kozlowski's account at a 10% interest rate, presumably the balance of Kozlowski's NQDC account at the beginning of 2000 was approximately $5.3 million ($219,543 divided by 4.1% = $5,354,707).

79. *Fannie Mae's Golden Eggs*, THE NEW YORK POST, Jan. 2, 2005, at 24 (stating that Raines received "a golden-parachute package of $26 million plus a *monthly* pension of $116,300" (emphasis added)).

80. *Id.*

81. Elizabeth Schwinn et al., *Nonprofit CEO's See Salaries Rise*, CHRONICLE OF PHILANTHROPY, Oct. 2, 2003, at 27.

82. Pacelle, *supra* note 25, at C1.

- Vice President Dick Cheney was criticized for receiving $2 million in bonus and NQDC from his former employer, Halliburton, after he claimed to have eliminated all financial ties and interests.[83]

After learning of the prevalence of excessive executive compensation, the public has demanded action. In 2003, another legislative proposal was issued in response to the "public outcry over outrageous deferred compensation plans for top executives as exposed by the Enron debacle and the current debate over the huge salary of the head of the New York Stock Exchange."[84]

C. *The Enron Compensation Report in General—Reflecting the Fundamental Compensation Problems*

In February 2002, at the request of the head of the Senate Finance Committee, the Joint Tax Committee launched a comprehensive study of Enron's tax and compensation practices. The result of these studies was the Enron Compensation Report, which describes many aspects of executive compensation at Enron and includes information on base salaries, bonuses, stock options, restricted stock, qualified plans, and NQDC.[85] While the Enron Compensation Report inspired the formation of new rules for NQDC embodied in Section 409A,[86] the facts in the report clearly reflect the Fundamental Compensation Problems, which are much more important than NQDC alone.

1. The Total Compensation Paid to Top Executives at Enron Was Grossly Excessive but NQDC Represented Less than 5% of Total Compensation

The Enron Compensation Report acknowledges that the media focused "on the magnitude of compensation paid to certain executives,"[87] and reported on

---

83. *CNN Live Today: People Along Gulf Coast Beginning Cleanup in Wake of Ivan; Kerry Speaks in New Mexico; Emmys on Sunday Night* (CNN television broadcast, Sept. 17, 2004) ("[T]he Independent Congressional Research Service found that under federal ethics laws, Dick Cheney did have a lingering financial interest in Halliburton.").

84. *Senate Panel Ban on COLI/BOLI Is Product of Intra-Industry Miscalculation*, INS. CHRON., Sept. 22, 2003, at 1, 6. The proposal would have required corporations to pay income tax on the death benefits received under a life insurance policy insuring an employee, if the employee had not worked for the corporation for a year or more. *Id.* at 6. Corporations frequently purchase life insurance to informally fund their obligations to pay employees under NQDC arrangements. *See* MICHAEL G. GOLDSTEIN ET AL., TAXATION AND FUNDING OF NONQUALIFIED DEFERRED COMPENSATION: A COMPLETE GUIDE TO DESIGN AND IMPLEMENTATION 180 (1998).

85. Enron Compensation Report, *supra* note 2, at 13-14.

86. Skonieczny & Pancoast, *supra* note 59 ("Congress has recently introduced several proposals to change the tax treatment of deferred compensation arrangements. . . . [T]hese proposed changes stem in part from the recent corporate scandals and a related public perception of excessive executive pay.").

87. Enron Compensation Report, *supra* note 2, at 13 ("Enron had a pay-for-performance

some rather incredible compensation figures at Enron. The compensation of the 200 most highly-compensated employees at Enron (the "Enron Top 200") was truly staggering. The compensation for the Enron Top 200 increased significantly in the years leading up to Enron's bankruptcy. Officer compensation doubled from 1998 to 1999 and tripled from 1999 to 2000.[88] In 2000, the year before Enron went bankrupt, each of the Enron Top 200 was paid over $1 million; at least 26 executives were each paid over $26 million; and three were paid over $100 million each, with the highest paid executive receiving $169 million.[89] The total compensation paid to the Enron Top 200 in 2000 was $1.4 billion, averaging $7 million per executive![90] Although Enron

---

compensation philosophy. Employees who performed well were compensated well."); *see also id.* at 14 ("Enron's long-term incentive program was designed to tie executive performance directly to the creation of shareholder wealth."); *id.* at 36 ("Enron's stated philosophy was a pay for performance approach to compensation . . . ."). Since Enron paid staggering amounts in bonuses immediately before declaring bankruptcy, the argument that all the contingent compensation paid at Enron was justified seems dubious. *Id.* at 14.

88. *Id.* at 545. The Enron Compensation Report noted that "in 2000 the deduction for compensation of officers was almost twice the deduction for salaries and wages." *Id.* The following table, adapted from the Enron Compensation Report, reflects the Enron Compensation Deductions for 1998, 1999, and 2000.

ENRON COMPENSATION DEDUCTIONS FOR 1998, 1999, AND 2000 (in millions of US $)			
	1998	1999	2000
Compensation of Officers	$149.9	$313.3	$952.5
Salaries and Wages	$499.7	$702.7	$557.5
Pension, profit sharing, etc., plans	$ 0.6	$ 0.8	$ 0.02
Employee benefit program	$344.7	$569.3	$1,456.8
Total	$994.9	$1,586.1	$2,966.8

Id. at 546. The phrase "employee benefit program" includes "contributions to employee benefit programs not claimed elsewhere on the [Enron tax] return (*e.g.*, insurance, health and welfare programs) that are not an incidental part of a pension, profit-sharing, etc., plan deducted in the previous line." *Id.* at 546 n.1623.

89. *Id.* at 561.

90. *Id.* at 13, 36, 562. The following chart, adapted from the Enron Compensation Report, reflects the compensation for the Enron Top 200 from 1998 through 2000.

went bankrupt in 2001, making it a horrible year for the company, the Enron Top 200 had a great year. Each member of the "club" again received over $1 million in compensation; at least 15 executives received over $10 million each; and one executive received over $56 million.[91] Further, at least 48 executives received bonuses of $1 million or more that year, and one executive received an $8 million bonus.[92]

NQDC certainly had its place in the Enron compensation structure. One statistic is remarkable—Ken Lay deferred $32 million under the Enron NQDC plan in one year.[93] NQDC was, however, less than 5% of the total

Compensation Paid to the Top 200 Highly-Compensated Employees For 1998 – 2000 (in millions of US $)					
Year	Bonus	Stock Options	Restricted Stock	Wages	Total
1998	$ 41.2	$ 62.0	$ 24.0	$ 66.1	$ 193.3
1999	$ 51.2	$ 244.6	$ 21.9	$ 84.1	$ 401.8
2000	$ 56.6	$ 1,063.5	$ 131.0	$ 172.6	$ 1,424.4
Total	$ 149.0	$ 1,370.1	$ 177.6	$ 322.8	$ 2,019.5

Id. at 547. From 1998 to 2000, the compensation to the top 200 highly-compensated employees increased drastically. "[Enron's] tax return information demonstrates that Enron's stock option deduction dramatically increased . . . . The deduction in 2000 was more than 1,000 percent greater than the deduction taken just two years prior." *Id.*; *see also id.* at 14 ("Enron's deduction for compensation attributable to the exercise of nonqualified stock options increased by more than 1,000 percent from 1998 to 2000."). Further, "stock option income resulting from the exercises of nonstatutory stock options for the top 200 most highly compensated employees was $62 million for 1998, $244.6 million for 1999, and $1,063.5 million for 2000." *Id.* at 547 n.1627. In other words, the 2000 total exceeded $1 billion. Incredibly, the Enron Compensation Report makes absolutely no recommendation for reform involving stock-based compensation. *Id.* at 41 ("In implementing its stock-based compensation programs, Enron appeared generally to follow IRS published guidance. Thus, no recommendations are made with respect to such programs.").

91. *Id.* at 561.

92. *Id.* at 566.

93. *Id.* at 14. Ken Lay was also provided with a $7.5 million line of credit, while Jeff Skilling borrowed $4 million from Enron in 1997, half of which was repaid in 1999, and the other half was repaid in 2001. *Id.* at 15. Also, "Enron entered into split-dollar life insurance arrangements with Mr. Lay ($30 million and $11.9 million), Mr. Skilling ($8 million), and Mr. Clifford Baxter ($5 million)." *Id.* With the enactment of the Sarbanes-Oxley Act of 2002,

compensation for the Enron Top 200 from 1998 through 2000.[94] The amount of NQDC at Enron was less than half of the amount paid as restricted stock, and was dwarfed by the amount Enron paid under its stock option programs. The yearly average deferral for all employees who participated in Enron's Deferral Plan was $30.5 million,[95] excluding Ken Lay's $32 million deferral in 2000.[96] In 2000, Enron paid the top 200 "$56.6 million of bonuses, $1.06

Congress has acted to curb compensation abuses involving loans; loans to top executives at public companies generally are now prohibited under Sarbanes-Oxley. *See* 15 U.S.C. § 78m(k) (2000).

94. The Enron Compensation Report states,

> In 1998, the 200 highest paid employees at Enron . . . deferred $13.3 million. By 2000, that amount had risen to $70 million. For the years 1998 through 2001, a total of $154 million in compensation was deferred. According to documents provided by Enron, Mr. Lay deferred $32 million under one of Enron's nonqualified deferred compensation plans.

Enron Compensation Report, *supra* note 2, at 14. Specifically, the total deferrals for the Enron Top 200 were:

1998	$13.3 million
1999	$19.7 million
2000	$67.0 million
2001	$54.4 million

Id. at 603. The percentage of total compensation that was deferred by the Enron Top 200 (under the NQDC plans) is estimated by the following chart:

	NQDC as a Percentage of Total Compensation for the Enron Top 200, 1998-2000			
	Total Compensation (Other than NQDC)*	Total Annual Deferrals Under the NQDC Plans	Total Compensation (including NQDC)	NQDC as a Percentage of Total Compensation
1998	$193.3 million	$13.3 million	$206.6 million	6.44 %
1999	$ 401.8 million	$19.7 million	$421.5 million	4.68 %
2000	$1,424.4 million	$67.0 million	$1,491.4 million	4.45 %
Total	$2,019.5 million	$100.0 million	$2,119.5 million	4.72 %

* *See supra* note 90.

95. $154 million - $32 million = $122 million/4 = $30.5 million per year.

96. Enron Compensation Report, *supra* note 2, at 604 n.1817.

billion of stock options, [and] $131.7 million attributable to restricted stock . . . ."[97]

Between 1998 and 2001, Enron essentially paid no income tax as a result of its enormous net operating losses.[98] Thus, Enron presented an environment particularly conducive to NQDC, which may have resulted in a higher percentage of compensation deferred at Enron than at most other corporations. A corporation's willingness to allow employees to defer compensation is usually tempered by its inability to claim an income tax deduction for NQDC until the compensation is actually paid to the employee.[99] An employer may prefer to pay the compensation immediately and obtain the income tax deduction, while the employee may want to use NQDC to delay her income tax liability.[100] Congress' Joint Committee on Taxation has described this situation as the "usual tension" between the employee and the employer over NQDC.[101] Because Enron had operating losses and was not required to pay taxes, the "usual tension" that typically restricts a corporation's willingness to allow an employee to defer compensation was absent at Enron.[102] But this perfect situation for NQDC typically will not exist. Thus, NQDC likely represents an even smaller percentage of total executive compensation at most companies.

2. Failure of the Enron Board of Directors to Engage in Arm's-Length Negotiations Over Executive Compensation

The Enron Compensation Report reveals a complete absence of oversight of executive compensation by the Board of Directors.

> This Report's detailed review of Enron's compensation programs reveals a process which *rested approval of executive compensation packages almost entirely with internal management.* Although the Compensation Committee of the Board of Directors formally approved both the total amount of

---

97. *Id.* at 13-14 (emphasis added).

98. Although Enron had financial statement income of $2.3 billion from 1996 through 1999, it paid no income taxes during those years (and actually had net operating losses for income tax purposes). Enron Compensation Report, *supra* note 2, at 5-6.

99. I.R.C. § 404(a)(11) (2005); *see* GOLDSTEIN, *supra* note 84, at 24 ("[N]o deduction for compensation under a NQDC plan is permitted until the benefit is taxable to the employee, regardless of whether the employer is a cash basis or accrual basis taxpayer.").

100. Some employers, however, may prefer to delay the payment to improve the corporation's cash position.

101. STAFF OF THE JOINT COMMITTEE ON TAXATION, 99TH CONG., GENERAL EXPLANATION OF THE TAX REFORM ACT OF 1986, 654 (Comm. Print 1987) [hereinafter the "1986 BLUE BOOK"]; *see also* GOLDSTEIN, *supra* note 84, at 189 (noting that the drafters of the Tax Reform Act of 1986 considered this "usual tension" as a rationale for the Act).

102. "Enron demonstrates that the theoretical tension between the employer's interest in a current tax deduction and the employee's interest in deferring tax . . . has little, if any, effect on the amount of compensation deferred by executives . . . because of [Enron's] net operating loss carryovers . . . ." Enron Compensation Report, *supra* note 2, at 634.

> compensation paid to executives and the form of such compensation, the Committee's approval generally was a *rubber stamp* of recommendations made by Enron's management.[103]

Although Enron's Compensation Committee hired outside consultant Towers Perrin,[104] as well as other consulting firms, to prepare studies evaluating the executive compensation arrangements, the Joint Tax Committee stated that "in some cases, the studies appeared to be designed to justify whatever compensation arrangement management wanted to adopt."[105] Other findings regarding executive compensation oversight include:

- "There was no indication that Enron's Compensation Committee ever rejected a special executive compensation arrangement brought to them."[106]
- "In [some] cases, the Compensation Committee either never reviewed certain arrangements for executives, or performed such a cursory review that they were not fully aware of what they were approving."[107]
- "The Compensation Committee did not scrutinize proposed arrangements, but basically approved whatever compensation arrangements were presented to them by management."[108]

In summary, the Enron Compensation Report concludes that "[t]he lack of scrutiny of compensation was particularly prevalent with respect to Enron's top executives, who essentially wrote their own compensation packages."[109]

D. NQDC at Enron Reflects the Fundamental Compensation Problems

NQDC was a part of the executive compensation package at Enron and, not surprisingly, the fundamental compensation problems of excessive

---

103. *Id.* at 19 (emphasis added).

104. *Id.* at 547, 554-57. The reliance on Towers Perrin was extensive.

> From Joint Committee staff interviews with many former members of the Compensation Committee, it appears that many members made decisions relying on the opinions of consultants without fully understanding the underlying issue. For example, former Compensation Committee members interviewed by Joint Committee staff could not explain why Enron purchased two annuities from Mr. Lay and his wife in 2001, but knew that Towers Perrin issued an opinion providing justification for the transaction.

Id. at 557.

105. *Id.* at 36.

106. *Id.*

107. *Id.* ("For example, a former chairman of the Compensation Committee could not remember an arrangement under which an Enron executive was awarded a fractional interest in an airplane as a form of compensation.").

108. *Id.* at 553.

109. *Id.* at 36.

compensation and lack of arm's-length bargaining by the Board of Directors were reflected in its NQDC program.

The headline grabber for Enron's NQDC program is that Ken Lay deferred $32 million in 2000.[110] In regards to the NQDC program generally, the Enron Compensation Report states, "Nonqualified deferred compensation was a major component of executive compensation for Enron. In 1998, the 200 highest paid employees at Enron . . . deferred $13.3 million. By 2000, that amount had risen to $70 million. For the years 1998 through 2001, a total of $154 million in compensation was deferred."[111] NQDC, however, represented less than 5% of total compensation to the Enron Top 200,[112] and was dwarfed by stock option compensation.[113]

As an indication of the Board's lack of arm's-length bargaining (or even oversight), the Enron Compensation Report reveals that "even though changes to the nonqualified deferred compensation plans were approved by the Compensation Committee, one former member of the Compensation Committee . . . did not know whether Enron offered nonqualified deferred compensation."[114] A separate "Deferral Plan Committee" of no more than three people was to be established to administer the NQDC Plan[115] and was authorized "to make, amend, interpret and enforce all appropriate rules and regulations for the administration of the . . . Deferral Plan and decide or resolve any and all questions, including interpretations of the . . . Deferral Plan."[116] Kenneth Lay, as the Enron CEO, was authorized to appoint the members of the Deferral Plan Committee.[117] Although Enron established its NQDC Plan in 1994, Ken Lay did not formally appoint the members of the Committee[118] until October 26, 2001 (less than two months before Enron filed for bankruptcy), when he appointed one person—Lawrence Gregory ("Greg") Whalley.[119]

E. The Report Asserts that the Enron Experience Reveals a Compensation Abuse Unique to NQDC—The Executives Can Control the Timing of the Payments After the Amount Has Been Deferred

Under the terms of Enron's NQDC plan, employees earning in excess of $120,000 could defer up to 35% of their salary, as well as up to 100% of bonuses and certain stock-based compensation.[120] Employees could defer as

110. *Id.* at 14, 604 n.1817.
111. *Id.* at 14.
112. *See supra* note 94.
113. *See supra* notes 90 and 94 and accompanying text.
114. Enron Compensation Report, *supra* note 2, at 553.
115. *Id.* at 605 n.1827, 611.
116. *Id.* at 611.
117. *Id.*
118. Apparently "informal" committees were created whenever issues arose. *Id.*
119. *Id.*
120. *Id.* at 606. In 1999, the floor salary for eligible employees was $130,000. But the

little as $2,000 per year.[121] Enron could also make supplemental contributions to a participant's NQDC account balance.[122] Approximately 300 employees participated in the 1994 Deferral Plan,[123] and as of December 2000 the deferred account balances totaled approximately $153.4 million.[124]

When an employee elected to defer income,[125] that employee could choose to receive payments in a lump sum or in up to fifteen annual installments upon termination of employment, including termination due to death or disability.[126] An employee could change the payment stream that would begin upon termination of employment if such election was made at least one year prior to the termination of employment.[127] For example, the employee could switch from an installment payment plan to a lump sum, or vice versa.

Enron's NQDC plan contained certain "accelerated distribution" provisions, which complicated the issue of constructive receipt and triggered great interest from the Joint Tax Committee.[128] A participant in the 1994 Deferral Plan could receive an immediate distribution of part or all of her deferred compensation account balance with the corporation's consent (the "Corporate Consent Requirement"), provided that the amount distributed, and the employee's account balance, would be reduced by 10% of the amount

---

floor was dropped to $120,000 in 2000. *Id.*

121. *Id.* The Enron NQDC plans included the 1994 Deferral Plan, the Enron Corp. Bonus Stock Option Program, the Enron Corp. Bonus Phantom Stock Program, and the 1998 Enron Expat Services, Inc. Deferral Plan. *Id.* at 568, 603.

122. *Id.* at 610 (stating that Enron contributed $500,000 per year to one executive's NQDC account). These are referred to as "supplemental" contributions because they do not reduce the amount of compensation otherwise payable to the executive.

123. From 1999 to 2001, there were approximately 340 participants, and as of December 2000 there were approximately 295 participants. *Id.* at 604.

124. *Id.*

125. The election to defer was irrevocable and had to be made in the year before the compensation was earned. *Id.* at 606.

126. *Id.* at 608.

127. *Id.* ("Payment elections . . . would not be effective until one full calendar year after receipt of the revised payment election form."). Although the ability of an employee to make a unilateral "subsequent election" was not specifically criticized in the Enron Compensation Report, the regulation of deferral extensions is an extremely important feature of Section 409A. *See infra* Part IV.C.3. While the Enron plan required the employee to make the election one year before terminating employment, the American Bar Association has argued that an employee should be entitled to make this type of unilateral election six months before terminating employment. American Bar Association Section of Taxation, Committee on Employee Benefits, *Report on the Proposed Restatement of the Service's Procedural Guidance for the Issuance of Advance Rulings on the Application of the Doctrine of Constructive Receipt to Nonqualified Deferred Compensation Arrangements*, 50 TAX LAW. 217, 235-36 (1996) [hereinafter *ABA Report*]. The ABA's argument was in response to a request from the IRS for comments on when an advance letter ruling could be issued for a NQDC plan under Revenue Procedure 71-19, 1971-1 C.B. 698, as amplified by Revenue Procedure 92-65, 1992-2 C.B. 428. *ABA Report, supra,* at 217.

128. Enron Compensation Report, *supra* note 2, at 608.

requested.[129] This 10% forfeiture provision frequently is referred to as a "Haircut Provision."[130]

A common question raised in any deferral plan is whether an employee has so much control over the money that she should be taxed immediately as it is earned, or whether the employee's income tax liability should be deferred. A cash basis taxpayer is generally subject to income tax on money actually received, or on money constructively received. As discussed in more detail below, an individual has "constructive receipt" of an amount if she has the unrestricted right to withdraw the money without significant restrictions or limitations.[131]

In *Metcalfe v. Commissioner*[132] the Tax Court held that a Corporate Consent Requirement in a NQDC plan is sufficient to prevent immediate taxation.[133] Based on this holding, one may argue that under the Enron deferral plan, the employees did not have constructive receipt notwithstanding the "accelerated distribution" provision. As such, it would be inappropriate to tax an employee because the corporation "might" agree to a prepayment. In the same way, it would be inappropriate to tax a seller of property on the total purchase price when the sales agreement calls for payments over ten years and grants the buyer the option, but not the obligation, to pre-pay.[134]

---

129. *Id.*

130. *See ABA Report*, *supra* note 127, at 235. Such a participant would not be permitted to participate in the plan for 36 months after the distribution. Enron Compensation Report, *supra* note 2, at 609. "The accelerated distribution provision was added to the plan in the First Amendment to the 1994 Deferral Plan (as restated effective August 11, 1997) dated October 13, 1997." *Id.* at 608. Prior to Section 409A's enactment, this technique was described as follows:

> *Haircut Provisions.* Many executive plans include "haircut" provisions, allowing employees to accelerate distribution of their benefits if they forfeit a specified percentage of their executive plan benefit. However, making "haircut" elections, or adding "haircut" provisions to an executive plan within the preference period before bankruptcy [generally 1 year for "insiders"], may not be sufficient to protect the executive plan assets in bankruptcy.

David N. Levine, *Is Chapter 11 Bankruptcy a Possibility? Have You Remembered Your Employee Benefit Plans*, PENS. & BEN. WEEK NEWSLETTER (RIA), Aug. 11, 2003; *see also* Schultz & Francis, *supra* note 68, at A1 ("Many companies do impose penalties, also known as haircuts, of 10% or in the case of Verizon and some other companies 6%.").

131. Treas. Reg. § 1.451-2(a) (2005); *see infra* notes 187-207 and accompanying text (discussing the details of the "constructive receipt doctrine").

132. 43 T.C.M. (CCH) 1393 (1982).

133. *Id.* at 1394, 1396, 1399.

134. Appropriately, the IRS has ruled that a Corporate Consent Requirement will not be respected if the participant is the majority shareholder. The rationale is that if the only person standing between you and the cash is you, you effectively have the ability to withdraw. I.R.S. Tech. Adv. Mem. 88-28-004 (Apr. 12, 1988). Accordingly, in that situation, the employee should be deemed to be in "constructive receipt" of the deferred amounts, and should be taxed on the compensation when it is earned, even if the amount will not be paid for many years.

The Enron deferral plan contained a provision that an employee would be prohibited from future participation in the plan for three years after requesting a withdrawal.[135] Such a provision could be considered a substantial limitation or restriction under the constructive receipt doctrine. The Enron Compensation Report specifically states that "[t]he plan was presumably designed this way to attempt to avoid constructive receipt."[136]

In December of 2000, one year before Enron declared bankruptcy,[137] of the 295 participants in the 1994 Deferral Plan, 181 participants requested accelerated distributions.[138] The Deferral Plan Committee, consisting of only Greg Whaley, approved 109 requests.[139] Approximately one-third of the deferred account balances were distributed.[140] The Enron Compensation

---

135. Enron Compensation Report, *supra* note 2, at 622. The argument would be that the employee must surrender a potentially valuable right, the right to participate in the NQDC plan for the next three years, to receive the distribution.

136. *Id.* at 622. A weaker argument also was available under the Enron 1994 Deferral Plan: because of the "haircut" provision under that plan, if an employee requested a $1,000 accelerated payment, she would receive $900 and forfeit $100. But Section 409A has rejected the argument that the "haircut" provision should prevent immediate taxation. *See infra* notes 297-98 and accompanying text. The Treasury Regulations provide that the existence of an early withdrawal penalty is *not* a substantial limitation or restriction for purposes of the constructive receipt doctrine. Treas. Reg. § 1.451-2(a)(2) (2005); *see also* Leonard G. Weld & Charles E. Price, *Constructive Receipt and the Substantial Restrictions Limitation*, 81 TAXES 45, 46 (2003) (summarizing several limitations and restrictions that are not considered substantial).

137. EmployeeCommittee.com: Deferred Compensation, http://www.employeecommittee.com/sr-deferredcomp_back.asp (last visited Nov. 11, 2005).

138. Enron Compensation Report, *supra* note 2, at 604, 622. There were no requests for accelerated distributions in 1998, 1999 or 2000. *Id.* at 622. There was a combined total of 350 participants in the 1994 Deferral Plan and the Expat Deferral Plan and 211 requested accelerated distributions in 2001. *Id.* at 622 & n.1883.

139. *Id.* at 611, 624. Mr. Whaley stated that because Enron had cash flow problems, he evaluated requests for early distributions in the same manner as requests for payments by other "unsecured creditors of Enron." *Id.* at 623. Specifically, "Enron would pay distribution requests made by active employees only, because active employees were needed to keep Enron operating, while inactive participants were providing no current service to Enron." *Id.* This seems like a typical business strategy—if the company has a cash flow problem, pay the electric bill in order to keep the lights on, and delay paying other bills if the creditor cannot jeopardize the company's operations. If Enron was "considered a going concern," then "all requests for early distributions would be approved." *Id.* Under the Expat Deferral Plan, 18 of 30 requests were approved, and distributions totaling $6.9 million were made. *Id.* at 624-25.

140. Under the 1994 Deferral Plan, approximately $46.2 million was distributed in response to requests for accelerated distributions and as of December 2000, the deferred account balances totaled $153.4 million. *Id.* at 604, 624. The figures provided by the Enron "Employee Committee" established to monitor the bankruptcy proceeding were slightly different.

> Between October 25, 2001 and November 30, 2001, approximately 206 plan participants made requests for accelerated distributions. Of this total, 126 requests were approved. This resulted in the distribution of approximately $54 million within 35 days of the bankruptcy filing.

Report asserts that these accelerated distributions prove that employees had too much "security"[141] and "control"[142] over deferred amounts. As a result, the Report concluded that the following fundamental changes to the NQDC rules were needed: (i) prohibit accelerated distributions;[143](ii) prohibit extensions or subsequent elections;[144] and (iii) prohibit distributions upon the occurrence of certain events.[145]

Under the Joint Tax Committee's recommendation to prohibit accelerated distributions, the employer would be prohibited from prepaying any deferred compensation, regardless of an economically motivated agreement subsequently reached between the parties. This prohibition would be comparable to proposing a tax rule that (i) any time A borrows money from B with payments to be made over a series of years, A should be prohibited from prepaying any of the principal, or (ii) any time A purchases property from B with a promissory note calling for payments over a series of years, A should be prohibited from making any prepayments.

The Joint Tax Committee's recommended prohibition of extensions eliminates economic flexibility in the opposite direction—once the employer and employee agree on when the NQDC will be paid, no extensions for payment can be granted. This is the equivalent of saying that a creditor cannot give a debtor an extension, but instead must immediately foreclose if the debtor is late in making a single payment.[146]

The Joint Tax Committee's third recommendation was to prohibit distributions upon the occurrence of certain events except in the event of "financial hardship, death, disability, retirement, the passage of a period of time specified . . . (e.g., three years), and change of control."[147] Although the Joint Tax Committee recommended an exception for "financial hardship," that exception is basically meaningless, as a practical matter, because the conditions

---

See EmployeeCommittee.com: Deferred Compensation, Update-July 23, 2004 www.employeecommittee.com/sr-deferredcomp.asp (last visited Nov. 11, 2005) [hereinafter Update July 23, 2004].

141. Enron Compensation Report, *supra* note 2, at 634.

142. *Id.* at 635; *see also id.* at 20 ("Changes should be made to the rules relating to nonqualified deferred compensation arrangements to curb current practices that allow for the deferral of tax on compensation income while providing executives with *inappropriate* levels of security, control, and flexibility with respect to deferred compensation." (emphasis added)).

143. *Id.* at 636.

144. *Id.* at 636-37. The Enron Compensation Report, however, does state, "Alternatively, limited opportunities to change elections could be provided . . . [but] the time that such elections are allowed to be made should be specified." *Id.* at 637.

145. *Id.* at 636.

146. Under this approach, a key issue is whether an employee will be subject to tax penalties if her employer has cash flow difficulties and fails to make a NQDC payment when scheduled and the employee fails to take immediate collection activities.

147. *Id.* at 636.

for obtaining a distribution under a NQDC plan due to a financial hardship are incredibly stringent.[148]

Thus, the Joint Tax Committee's three proposed changes would eliminate flexibility after the employer and employee agree to a deferral, even if economic conditions require changes.

F. The NQDC Problem Asserted by the Enron Compensation Report Does Not Exist

The Enron Compensation Report asserts that the "accelerated distribution" provisions of the Enron plan provided too much "flexibility," "security," and "control" over the NQDC because it allowed the Enron employees to receive $46.2 million in cash just before Enron declared bankruptcy, to the disadvantage of other Enron creditors.[149] There are several fundamental flaws with the Joint Tax Committee's analysis, three of which are discussed below.

1. Under the Bankruptcy Laws, All of the Accelerated Distributions to Enron's Top Executives in the NQDC Plan Can Generally Be Recovered by the Bankruptcy Trustee and Distributed to Enron's Creditors

If an employer transfers property within a certain period of time before declaring bankruptcy (the "look-back period"), the bankruptcy trustee can recover the property and distribute it among the employer's creditors under the priority rules established by the bankruptcy laws. Generally, the look-back period for transfers is ninety days, but if the transfer was made to an "insider" (e.g., an officer, director, or substantial stockholder), the look-back period is one year.[150]

The application of bankruptcy priority rules to NQDC plans was demonstrated in the case of *In re Bank Building*.[151] In this case, a corporate director deferred his director's fees for several years, and the entire deferred account balance was distributed to the director less than a year before the company declared bankruptcy under Chapter 11.[152] The court held that although the NQDC account was established several years before the corporation declared bankruptcy, the relevant transfer of funds occurred during

148. Rev. Proc. 92-65, 1992-2 C.B. 428; *see also* Enron Compensation Report, *supra* note 2, at 625 ("There were 11 [hardship] requests from the 1998 Deferral Plan, one request from the 1992 Deferral Plan, and three requests from the Project Participation Plan. . . . [N]o hardship requests were granted. . . . Documents provided by Enron show that one hardship request was granted from the 1992 Deferral Plan in 1998.").

149. *See supra* notes 140-42 and accompanying text.

150. *See* 11 U.S.C. § 547(b)(4) (2000); *see also id.* § 101(31) (defining the term "insider").

151. *In re* Bank Building & Equip. Corp. of Am., 158 B.R. 138 (E.D. Mo. 1993).

152. *Id.* at 139.

the look-back period; thus, the bankruptcy trustee was entitled to recover the payment and redistribute it among the corporation's creditors.[153]

As of July 23, 2004, the Enron Bankruptcy Trustee had already recovered approximately $11 million of the $53 million paid in accelerated distributions under the Enron NQDC plans[154] and is still in the process of collecting remaining amounts.[155] The website for the Enron "Compensation Committee" states, "These funds . . . will be distributed to the unsecured creditors in various classes on a pro rata basis based on the amount of an allowed claim."[156]

---

153. *Id.* at 140.

154. *See* Update July 23, 2004, *supra* note 140 ("To date [as of July 23, 2004], the Debtors and the Employee Committee have successfully collected approximately $11 million of the $53 million in deferred compensation.").

155. As stated on the website for the Enron employees:

> [T]he accelerated distributions are subject to avoidance actions, which means that the distributions were *preferences* to some creditors to the detriment of other creditors who were not given a distribution. The Debtors, the Creditors Committee and the Employee Committee agreed that the funds distributed from the plans should be recovered.

EmployeeCommittee.com: Deferred Compensation, Enron and the Employee Committee Discussions, www.employeecommittee.com/sr-deferredcomp_disc.asp (last visited Nov. 11, 2005). In recovering the accelerated NQDC distributions, the following settlement strategy was proposed:

- Employees who either voluntarily terminated employment before the bankruptcy date (December 2, 2001) or were involuntarily terminated for cause [referred to as "Group III"] could repay 90% of the accelerated distribution received. *See* EmployeeCommittee.com: Deferred Compensation, Proposed Settlement www.employeeommmittee.com/sr-deferredcomp_prop.asp (last visited Nov. 5, 2005). Presumably this 10% concession was merely used to encourage prompt payment and avoid attorney fees and collection costs.
- Employees whose employment was terminated without cause after the bankruptcy filing [referred to as "Group II"] were asked to repay from 40% to 85% of the accelerated distribution amount by October 31, 2003, and if the amount was not repaid by that date, litigation would be commenced to collect 100%. *Id.*
- Active Enron employees [referred to as "Group I"] were required to repay 40% of the accelerated distribution amount by October 31, 2003, and if the amount was not repaid by that date, litigation would be commenced to collect 100%. *Id.*

The repayment terms offered were as follows:

(i) Cash repayment in a single lump sum;

(ii) Active employees . . . will have the repayment options of payment in installments or in a lump sum from their current net compensation. Full payment is required by October 31, 2004;

(iii) Former employees may elect to make the installment payments with payment in full no later than October 31, 2004.

EmploymentCommittee.com: Deferred Compensation, Repayment Terms www.employeecommittee.com/sr-deferredcomp_repay.asp (last visited Nov. 11, 2005).

156. *See* Update July 23, 2004, *supra* note 140.

2. Even With No NQDC, Bonuses Can Be Used to Provide Extra Compensation at Any Time

Enron paid $104.9 million in "11th Hour Bonuses" in the thirty-five days immediately prior to declaring bankruptcy, which was almost twice as much as it paid in accelerated NQDC distributions during that time.[157] A total of forty-eight Enron executives received a bonus of $1 million or more in 2001 and some individuals received bonuses in the range of $5 million to $8 million.[158] "Enron's bankruptcy filing . . . shows that bonuses to 144 insiders . . . paid during the year preceding the bankruptcy totaled approximately $97 million."[159] The Bonus Plan was executed on November 28, 2001, a mere five days before Enron filed bankruptcy and was implemented "to recognize, motivate, and reward exceptional accomplishment of annual corporation objectives during calendar year 2001."[160] As a further indication that Enron's Board of Directors and the Compensation Committee merely acted as a "rubber stamp" for the compensation decisions made by Enron's top executives, the Board approved a bonus plan that authorized management "to modify the list of employees and payment amounts as deemed appropriate."[161]

Although the bonus program was adopted just five days before Enron declared bankruptcy, Enron's law firm, Weil, Gotshal, and Manges, advised that the program "was a compensation design for which reasonable justification existed."[162] Enron's compensation consulting firm, Towers Perrin, argued that the "payments were consistent with industry practices."[163] Even though the Board of Directors, the Compensation Committee, the law firm, and the consultants approved the plan, the Enron executives themselves took some

---

157. Enron Compensation Report, *supra* note 2, at 573. The Report states,

> In the weeks immediately preceding the bankruptcy, Enron implemented bonus programs; one for approximately 60 key traders and one for approximately 500 employees that Enron claimed were critical for maintaining and operating Enron going forward. In order to receive a bonus under one of these programs, the employee had to agree to repay the bonus, plus an additional 25 percent, if the employee did not remain with Enron for 90 days. The combined cost of the programs was approximately $105 million.

Id. at 14. The Board of Directors approved the bonus plan for traders on November 18, 2001 and for "critical noncommercial staff (i.e., persons other than traders)" on November 28, 2001. *Id.* at 574. The 584 employees who participated in the two bonus plans received bonus payments in the range of $2,500 to $8 million. *Id.* at 577.

158. *Id.* at 566.

159. Enron Compensation Report, *supra* note 2, at 566. Under the federal bankruptcy laws, "insiders" include corporate officers, directors, and substantial shareholders. 11 U.S.C. § 101(31) (2000).

160. Enron Compensation Report, *supra* note 2, at 575.

161. *Id.*

162. *Id.* at 576. ("Weil, Gotshal, and Manges commented that it was not a legal decision to implement this type of plan, but that it was an issue of business judgment that could be second-guessed.").

163. *Id.*

unusual steps. The bonus payments "were made in cashier's checks (net of payroll taxes) and were paid on the Friday preceding Enron's bankruptcy filing." Some employees claimed that "there was an air of secrecy involving the payments," and some employees who heard about the bonuses "waited in the office for the payments."[164]

The same bankruptcy rules that allow the Enron Bankruptcy Trustee to recover the accelerated NQDC distributions also allow the Enron Bankruptcy Trustee to recover 11th Hour Bonuses. The Bankruptcy Trustee has reached settlements with some bonus recipients and is pursuing litigation against other bonus recipients.[165]

3. The Enron Example Fails to Demonstrate NQDC Abuses

Although the Enron Compensation Report asserts that the Enron situation demonstrates that employees have extensive control, security, and flexibility with a NQDC plan, the actual facts in the Enron situation fail to demonstrate that control, security, and flexibility are the problems. Instead, the compensation problems at Enron were excessive amounts of compensation and lack of oversight by the Enron Board of Directors and Compensation Committee. The accelerated distributions from the Enron NQDC plan were "preferences" under the Bankruptcy laws, and the Bankruptcy Trustee has recovered and is recovering the accelerated distributions, which will be distributed for the benefit of other Enron creditors. In addition, the Enron Board's approval of 11th Hour Bonuses vividly demonstrates the danger posed when the board of directors and its Compensation Committee are merely "rubber stamps" for management's compensation decisions that permit management to effectively trigger compensation payments at will through the use of a bonus program, regardless of whether the company uses NQDC.

In the end, the Enron NQDC plan actually provided a significant benefit to Enron's creditors, and caused the Enron executives to receive less compensation than they would have received in the absence of the NQDC plan. Specifically, in the absence of the NQDC plan, the Enron executives would have received greater compensation in the years when the compensation was earned. Because those compensation amounts were deferred, they were either (i) held by Enron when it declared bankruptcy, or (ii) received by the executives as accelerated distributions shortly before Enron declared bankruptcy. In either case, the Bankruptcy Trustee could recover the amounts and distribute them to the Enron creditors.

---

164. *Id.*

165. *See* EmployeeCommittee.com: 11th Hour Bonuses, Update: August 9, 2004, www.employeecommittee.com/sr-11hrbonus.asp (last visited Nov. 11, 2005) (a lawsuit was filed on December 1, 2003, against approximately 290 recipients of 11th Hour Bonuses).

III. THE ROLE OF NQDC IN GENERAL: "DON'T SHOW ME THE MONEY . . . YET!"

NQDC has likely existed since the dawn of civilization. In antiquity, many military leaders promised their soldiers that they would be highly compensated after their victory in exchange for dedicated service during the war.[166] In the Bible, Jacob worked fourteen years for Laban in exchange for permission to marry Laban's younger daughter, Rachel.[167] Parties often agree to payments of NQDC out of simple economic necessity—the service recipient (the employer) buys services on credit; the service provider (often referred to in this Article as the "employee") accepts an IOU for services rendered. Instead of adopting the motto "Show Me the Money!"[168] from the movie *Jerry Maguire*, with NQDC, the employee says "Pay Me Later."

The term "nonqualified" refers to the fact that these arrangements do not satisfy the complex rules of section 401 of the I.R.C. ("Section 401"), or the onerous requirements of ERISA.[169] They are referred to as "deferred compensation" arrangements because the service provider performs the services at the present time, but the service recipient need not pay until a future tax year.

---

166. This analogy is particularly appropriate because if the soldiers were defeated, they would forfeit all the promised future compensation. Similarly, an executive may never receive promised NQDC if the company goes bankrupt. *See supra* Part II.F.1.

167. *Genesis* 29:18-35.

168. JERRY MAGUIRE (TriStar Pictures 1996).

169. The Employee Retirement Income Security Act of 1974, 29 U.S.C. §§ 1001 to 1461 (2000) [hereinafter ERISA], and the IRC, generally require that qualified pension plans satisfy several requirements:

(i) The plan must not discriminate based on participation or benefits. *See* I.R.C. § 401(a)(4) (2005) (preventing discrimination in benefits as between highly compensated employees and lower paid employees); *id.* § 401(a)(26)(A) (requiring that a certain percentage of the employees participate in the plan).

(ii) The plan must satisfy certain vesting rules. *See* ERISA § 203(a), 29 U.S.C. § 1053; I.R.C. § 411(a)(12) (allowing an employer to delay vesting by using 5-year "cliff" vesting or 7-year "graded" vesting).

(iii) The plan must satisfy minimum funding and trust establishment guidelines. *See* ERISA § 302, 29 U.S.C. § 1082; ERISA § 403(a), (d), 29 U.S.C. § 1103(a).

(iv) The plan must abide by reporting and disclosure rules. *See* ERISA § 101, 29 U.S.C. § 1021; *see also* GOLDSTEIN, *supra* note 84, at 125-26 (stating that the plan administrator is also required to file IRS Form 5500 annually).

(v) The plan must establish a claims procedure. *See* 29 C.F.R. § 2560.503-1 (2005).

(vi) The plan must also impose fiduciary duties on the persons who exercise discretion or control of management of the funds. *See* ERISA § 404(a)(1), 29 U.S.C. § 1104(a)(1).

See generally GOLDSTEIN, *supra* note 84, at 123-26 (discussing requirements that are generally applicable to qualified plans).

A. Bona Fide Economic Reasons for Establishing NQDC Arrangements

A variety of non-tax reasons motivate parties to establish NQDC arrangements. First and foremost, NQDC effectively allows the employer to buy services on credit.[170] An employer may prefer not to pay all of its compensation expenses immediately in cash for a variety of reasons. For instance, fluctuations in the business cycle may suggest a natural NQDC arrangement. Examples include a farm owner who prefers to pay more compensation after harvest time, or a subcontractor who prefers to pay his employees at least a portion of their compensation when he is paid for completing the construction job. Another example is the importer who takes her employees on a buying trip overseas and pays the bulk of their compensation after they return home and sell a large quantity of the goods purchased on the trip. In some situations, the payment delay may be perfectly acceptable to the service provider for practical reasons. For example, a crew may be at sea for an extended period of time before returning to port, and the ship owner may prefer to delay compensation payments in the interest of preserving cash flow. The crew may be perfectly happy to delay receipt of payment until arriving in port because the crew can spend their compensation on a wide variety of goods and services in port that are unavailable when the ship is out to sea.

Parties may desire to base payments on the occurrence (or non-occurrence) of future events. In an especially famous NQDC arrangement, actor William Holden agreed to reduce his current compensation for working on the 1957 Academy-Award winning film, *The Bridge on the River Kwai*,[171] in exchange for future compensation payments equal to 10% of the film's gross profits.[172] Although the film was a huge success, Holden was unfortunate in that his NQDC agreement capped the amount he could receive at $50,000 per year.[173]

---

170. An alternative would be for the service recipient to provide equity-based compensation, such as stock options or phantom stock, to the service provider. The use of equity-based compensation can be attractive for the employer because it does not impact cash flow. But equity-based compensation dilutes the percentage interest of existing owners, thereby decreasing their total return if the enterprise is successful. *See* Thomas Z. Reicher et al., *Statutory Stock Options*, in 381-2d TAX MANAGEMENT PORTFOLIOS (BNA 2004); John Utz, *Nonstatutory Stock Options*, 383-2d TAX MANAGEMENT PORTFOLIOS (BNA 2001). *See generally* NEAL A. MANCOFF & DAVID M. WEINER, NONQUALIFIED DEFERRED COMPENSATION ARRANGEMENTS (1998) (discussing the use of equity-based compensation).

171. THE BRIDGE ON THE RIVER KWAI (Columbia Pictures 1957).

172. 3 BORIS BITTKER & LAWRENCE LOKKEN, FEDERAL TAXATION OF INCOME, ESTATES AND GIFTS ¶ 60.2.2 (rev. 3d ed. 2005).

173. One 1959 article stated:

> For appearing in *Bridge on the River Kwai*, William Holden agreed to 10 percent of the gross, but for tax reasons wanted it paid to him at the rate of only $50,000 a year. The picture has already made so much money (between $20 and $30 million) that Holden's share now stands at between $2,000,000 and $3,000,000. Not only will it take 40-year-old Holden at least 40 years to get the last of his money, but [the producer] can in the

Holden died long before receiving an amount even close to 10% of the gross profits.[174]

Second, NQDC can function as a savings vehicle for a service provider. The Federal Court of Claims has noted that "professional athletes structure their payment schedule to ensure a degree of financial stability" because the length of their professional careers is uncertain.[175] Sugar Ray Robinson is one such professional athlete who utilized NQDC as a mechanism for savings:

> In 1957, Sugar Ray was 36 years old, had been a professional prizefighter for over 20 years, had been in approximately 170 professional fights and had been world middleweight champion five times. A prizefight between Sugar Ray, the then world middleweight champion, and Carmen Basilio, the then world welterweight champion, was scheduled for September 23, 1957 at Yankee Stadium in New York City for the world middleweight championship. On July 31, 1957, Sugar Ray and the promoter entered into a contract providing that Sugar Ray would receive 45% of the ticket revenue and 45% of the radio, theater/television, and motion picture receipts. According to the agreement, Sugar Ray would receive 40% of his payment within two weeks of the fight, 20% in quarterly payments in 1958, 20% in quarterly payments in 1959, and 20% in quarterly payments in 1960. Thus, Sugar Ray used a NQDC arrangement to stretch out the payments from the fight over 4 years.[176]

Third, NQDC provides "teeth" to non-compete arrangements that would otherwise be unenforceable. Employers frequently impose non-compete provisions on key employees who possess information that, if disclosed, could be very damaging to the employer and very valuable to its competitors.[177] For example, the employer may have enabled the employee to gain valuable business customers and contacts during the course of employment, and the loss of those customers and contacts could be detrimental to the employer. Often the applicable state law renders non-compete agreements unenforceable after a relatively short period of time, usually no more than two or three years from the

---

meantime invest it and make well over $50,000 a year, thus in effect having got Holden's services in *Kwai* for nothing.

Id. (quoting TIME, Jan. 19, 1959, at 66).

174. William Holden (birth name, William Franklin Beedle, Jr.) died in 1981, 24 years after the release of *Bridge on the River Kwai*. *See* GOLDSTEIN, *supra* note 84, at 12.

175. Buffalo Bills, Inc. v. United States, 31 Fed. Cl. 794, 800 (1994) ("The length of a professional athlete's career depends upon his ability to out-perform others who are themselves aggressively competing for coveted employment contracts.").

176. GOLDSTEIN, *supra* note 84, at 12; *see* Robinson v. Comm'r, 44 T.C. 20, 22 (1965), *acq.*, 1970-2 C.B. xxiii, *acq.*, 1976-2 C.B. 2.

177. *See* 1 BRIAN M. MALSBERGER, COVENANTS NOT TO COMPETE, A STATE-BY-STATE SURVEY ix (Samuel M. Brock & Arnold H. Pedowitz, eds., 4th ed. 2003) ("Restrictions on postemployment activities designed to protect legitimate employer interests are relevant today more than ever.").

date the employee terminates employment.[178] In order to create an economic incentive for the employee to refrain from competing for a longer period of time, an employer may withhold a portion of the employee's compensation and will pay the amount only if the employee satisfies the extended non-compete period.

Fourth, businesses may enter into NQDC arrangements because of cash flow needs. A start-up company may have a great business plan, but may lack essential capital to pay accountants, attorneys, architects, and other necessary service providers. The company could issue stock or other equity interests to such service providers, but this will reduce the original owners' potential returns if the venture is successful.[179] As an alternative, the start-up could agree to pay the service providers at a later date when cash flow is expected to be available.

In addition, owner-employees of small businesses frequently agree to take below-market compensation in the early years to allow the business to grow, but expect to be paid above-market compensation in later years after the business is established. In one case, the Tax Court concluded that payments to a deceased employee's widow were tax-deductible as "reasonable compensation" because the employee was underpaid during his lifetime.[180]

Finally, NQDC benefits are useful as an employee retention device. An employer might offer an employee a specified sum per year after termination of employment if the employee remains a full-time employee until normal retirement age. As the employee approaches normal retirement age, the promised benefit may discourage the employee from quitting and going to work for a competitor. These arrangements frequently are called "golden handcuffs."[181]

B. The Income Tax Rules for NQDC Before Section 409A—Substantial Flexibility Over Time of Payment

In a NQDC arrangement, the parties agree that the employee will be paid in a future tax year for services rendered currently. Typically employees would like to delay paying income tax on benefits while maintaining maximum flexibility over the payment schedule. But the income tax rules have always restricted the employee's flexibility to some extent.

> *EXAMPLE #2.* Debra Deferral and her employer agree that it will *not* pay her $5,000 of the compensation she will earn during 2006 ($192.00 per pay period, for 26 pay periods), but instead will hold on to it. Those deferred amounts will be credited to a bookkeeping account the company maintains,

---

178. *See, e.g., id.* at 1211 (discussing Illinois cases in which two-year restrictions were held to be reasonable and enforceable).

179. *See supra* note 170.

180. Way Baking Co. v. Comm'r, 27 T.C.M. (CCH) 168, 172 (1968).

181. *See* GOLDSTEIN, *supra* note 84, at 13.

> and the account balance will grow at the prime rate of interest. Debra does not want to pay income tax on the $5,000 deferral until she actually receives the cash payment and would like to be able to demand payment of the benefit at any time. For example, if she experiences financial difficulty or wishes to make a major purchase, she wants to be able to demand and receive the benefit immediately. On the other hand, she wishes to retain the flexibility of allowing the benefit to accrue tax-free (for Debra)[182] until she needs the money at retirement.

Before Section 409A, this tension between income deferral and payment flexibility was addressed mainly by the "constructive receipt doctrine." This theory continues to apply even after the enactment of Section 409A.[183] The constructive receipt doctrine flows from the fundamental tax principle that a taxpayer should be obligated to pay tax only when she has the "ability to pay"—if a taxpayer has no cash, she has no ability to pay and should not be obligated to pay income tax.[184] Based on the ability to pay concept, a cash basis taxpayer is taxed on an amount when she (i) actually receives it, or (ii) *constructively* receives it.[185] A taxpayer "constructively" receives an amount only when she has an "unrestricted right to receive it immediately."[186] Under this definition, even before the enactment of Section 409A, an employee would not have been given an unrestricted right to withdraw her NQDC benefit because such a right would cause her to be taxed immediately on the benefit.

Although the employee could not escape immediate taxation, parties could still build flexibility into a NQDC arrangement. The applicable regulations provide that "income is not constructively received if the taxpayer's control of its receipt is subject to *substantial limitations or restrictions*."[187] If the amount is not available until the passage of some period of time, the taxpayer will not be deemed to have constructive receipt until the period of time has passed.[188] When designing a NQDC plan, the parties could specify certain events that would trigger NQDC payments to the employee or the employee's beneficiary,

---

182. Since the deferred amount is retained by the employer, the employer would pay tax each year on the taxable earnings of the deferred amounts. *See* GOLDSTEIN, *supra* note 84, at 33-34.

183. I.R.C. § 409A(c) (2005) ("Nothing in [§ 409A] shall be construed to prevent the inclusion of amounts in gross income under any other provision of this chapter . . . or any other rule of law earlier than the time provided in this section."); I.R.S. Notice 2005-1, 2005-2 I.R.B. 274 ("§ 409A does not alter or affect the application of any other provision of the Code or common law tax doctrine.").

184. *See* NEWMAN, *supra* note 41, at 15 ("Fairness in taxation is often discussed in terms of ability to pay. Those who are able to pay more taxes should pay more; those who are able to pay less taxes should pay less.").

185. Treas. Reg. § 1.446-1(c)(1)(i) (2005) (including the "cash receipts and disbursements method" as a permissible method of accounting).

186. Childs v. Comm'r, 103 T.C. 634, 654 (1994).

187. Treas. Reg. § 1.451-2(a) (2005) (emphasis added).

188. Rev. Rul. 60-31, 1960-1 C.B. 174, 180.

while still permitting the employee to claim that the NQDC benefit is subject to a "substantial limitation or restriction." The chart below outlines a few common "Triggering Events" that leave the employee's right to the NQDC benefit subject to "substantial limitations or restrictions."

Triggering Event	**Explanation of Why a "Substantial Limitation or Restriction" Exists**
Death	While some people might say they are "dying" to receive a benefit, few people would choose to die in order to trigger a benefit payment.
Disability	Few people would choose to become disabled to trigger a NQDC payment.
Termination of Employment	If a person quits his job to receive a NQDC benefit, he is forfeiting the right to receive future salary, bonuses, and other benefits connected with the job.

Over the years, one could say that the IRS and the courts went "trigger happy," approving at least 13 Triggering Events that could be used in a NQDC plan:

1. the employee "reaches age 60"[189]
2. the employee "becoming partially or totally incapacitated"[190]
3. the employee works "for a period of five years"[191]
4. the "termination of the [employee's] employment by the corporation"[192]
5. the full-time employee "becoming a part-time employee"[193]
6. the employee "ceases to be a director of the corporation and becomes a proprietor, officer, partner, employee, or otherwise

189. Rev. Rul. 60-31, 1960-1 C.B. 175. This list of 13 triggering events is set forth in GOLDSTEIN, *supra* note 84, at 58.

190. Rev. Rul. 60-31, 1960-1 C.B. 175.

191. *Id.*

192. *Id.* This Revenue Ruling expounded upon termination by providing that the employee would *not* forfeit his NQDC benefits if he "should fail or refuse to perform his duties." Rev.Rul. 60-31, 1960-1 C.B. 175. Thus, the employee could quit or be terminated for cause by his employer and still receive his NQDC. While this might be viewed as a "withdrawal right" because the employee can obtain the NQDC benefits at any time by refusing to work, the IRS concluded that losing one's job is a "substantial limitation or restriction" on the ability to withdraw the NQDC benefits, so that the employee would not be subject to immediate taxation on the NQDC benefits under the constructive receipt doctrine. *See* Treas. Reg. § 1.451-2(a) (2005).

193. Rev. Rul. 60-31, 1960-1 C.B. 175.

becomes affiliated with any business that is in competition with the corporation"[194]

7. the employee being entitled to receive payments "beginning with the first day of the calendar year immediately following the year in which the director ceases to be a director"[195]
8. the employee's employment terminates (whether because of "disability, retirement, death, discharge or *resignation*")[196]
9. the employee reaches age 65[197]
10. the employee "leaves the [corporation's] service area to [work] elsewhere"[198]
11. a "change of control" of the corporation[199]
12. a decrease of the corporation's net worth below $10 million[200] or
13. "the Employer is liquidated, pursuant to a transaction whereby no successor corporation assumes the assets and liabilities of the Employer."[201]

Triggering Events that meet the individual needs of the employee could also be designed. The Tax Court has stated that "[t]he doctrine of constructive receipt is based on the principle that income is received by cash method taxpayers '*when it is made subject to the will and control of the taxpayer . . .*'"[202] Presumably, as long as the occurrence of an individually designed Triggering Event was beyond the employee's control, it could be used in the NQDC arrangement. But ultimately courts will determine whether an

194. Rev. Rul. 71-419, 1971-2 C.B. 220.

195. *Id.*; *see* Robinson v. Comm'r, 44 T.C. 20, 25 (1965); *see also* Rev. Rul. 70-435, 1970-2 C.B. 101 (discussing the constructive receipt doctrine as it applies to plans that incrementally disburse payments).

196. Metcalfe v. Comm'r, 43 T.C.M. (CCH) 1393, 1395 (1982).

197. Goldsmith v. United States, 586 F.2d 810, 813 (Ct. Cl. 1978).

198. Minor v. United States, 772 F.2d 1472, 1473 (9th Cir. 1985).

199. I.R.S. Priv. Ltr. Rul. 95-08-014 (Nov. 22, 1994) ("The Plan provides that benefits will be paid upon . . . the voluntary termination of the plan by a corporate successor."); I.R.S. Priv. Ltr. Rul. 87-46-052 (Aug. 19, 1987) (stating that amounts are paid after an involuntary termination following a change of control); I.R.S. Priv. Ltr. Rul. 84-18-095 (Jan. 31, 1984) (stating that deferred amounts become immediately payable upon a change of control).

200. I.R.S. Priv. Ltr. Rul. 95-08-014 (Nov. 22, 1994) ("The Plan automatically terminates if the Company's net worth falls below $10,000,000."). A copy of the NQDC plan considered in I.R.S. Priv. Ltr. Rul. 95-08-014 is reprinted in 24 TAX MGMT. COMP. PLANNING J. 9 (Jan. 5, 1996). Article 7.3 of the plan document provides, "All other provisions of the plan notwithstanding, in the event the net worth of the Company falls below ten million U.S. dollars, ($10,000,000), the Plan will automatically terminate and all Participant Accounts shall be immediately distributed to the Participants." 24 TAX MGMT. COMP. PLANNING J., *supra.*

201. I.R.S. Priv. Ltr. Rul. 84-35-031 (May 24, 1984).

202. Furstenberg v. Comm'r, 83 T.C. 755, 791 (1984) (quoting Loose v. United States, 74 F.2d 147, 150 (8th Cir. 1934)) (emphasis added).

individual has constructive receipt based upon the particular facts and circumstances of the case.[203]

Before the enactment of Section 409A, a NQDC plan was more flexible because the plan could be designed so that the employee would receive the NQDC benefits upon the first occurrence of a variety of Triggering Events. Other key sources of flexibility before the implementation of Section 409A were (i) the ability of the employer and employee to mutually agree to *accelerate or extend* NQDC payments, and (ii) the employee's ability to unilaterally *extend* NQDC payments shortly before the occurrence of a Triggering Event. A series of court cases established, over IRS objections, that after an amount had been deferred, the parties subsequently could agree to either extend or accelerate payments, and the employee would not be subjected to income tax on the NQDC benefits until the employee actually received the cash.[204] The employees in these cases did not have constructive receipt because each employee's ability to accelerate the NQDC benefit was subject to the limitation or restriction that the employer may not agree to make the distribution.[205]

In *Metcalfe*, the IRS argued that because the employer had a policy of accelerating payments to any requesting employee, the employees had an unrestricted withdrawal right over the NQDC money, and under the constructive receipt doctrine, all the participants should be immediately taxed on all the accrued NQDC benefits.[206] The Tax Court concluded that the employees were not in constructive receipt of the NQDC benefits because the employer had the legal right to deny a request for an acceleration.[207] Thus, the Tax Court recognizes a key economic reality that Section 409A ignores: the need to obtain the approval of an adverse party is a "substantial limitation or restriction" upon the employee's ability to receive the NQDC payments.[208] Nevertheless, this technique would not be available if the employee is the majority shareholder of the company or otherwise controlled the employer.[209]

---

203. *See* Childs v. Comm'r, 103 T.C. 634, 654 (1994) ("The doctrine of constructive receipt is essentially a question of fact." (citation omitted)).

204. *See* Oates v. Comm'r, 18 T.C. 570, 584-85 (1952); Veit v. Comm'r, 8 T.C.M. (CCH) 919, 922 (1949) [hereinafter *Veit II*]; Veit v. Comm'r, 8 T.C. 809, 816, 818 (1947) [hereinafter *Veit I*].

205. The need to obtain the employer's consent can be viewed as a significant limitation or restriction because the employer presumably would have a natural tendency to deny the request, and not impair its cash flow. On the other hand, some employers might prefer to accelerate the income tax deduction by paying the NQDC early. I.R.C. § 404(a)(5) (2005).

206. Metcalfe v. Comm'r, 43 T.C.M. (CCH) 1393, 1398 (1982).

207. *Id.* at 1399.

208. *Id.* at 1398 ("The fact that [the employer], under some circumstances, might have been willing to rescind the payout provision and accelerate the payout does not give [the taxpayer] any unrestricted *right* to the withheld funds." (emphasis added)).

209. Rev. Proc. 96-3, 1996-1 C.B. 456, 456, 461 (stating that the IRS will not issue a favorable advance letter ruling regarding a NQDC arrangement for a controlling shareholder); I.R.S. Tech. Adv. Mem. 88-28-004 (Apr. 12, 1988) ("A taxpayer who is a controlling

In those situations, the requirement to obtain the employer's consent would not be a "substantial limitation or restriction."

Another popular technique to create additional flexibility in the NQDC plan was a provision that the employee could unilaterally elect to change the form or timing of the payments. For example, a NQDC plan might provide for a lump sum payment at normal retirement age, but allow the employee to unilaterally elect to receive installment payments over five, ten, or twenty years, or over her life expectancy, as long as the employee filed the election with the employer at least one year before attaining normal retirement age. The popularity of this technique was reflected when the American Bar Association requested that the IRS establish time limits for exercising this type of election.[210]

Thus, while an employee could not have had an unrestricted withdrawal right over NQDC amounts (and avoid immediate taxation), the employee deferring money under a NQDC arrangement could have enjoyed some flexibility regarding the timing of payments if (i) multiple Triggering Events were included in the NQDC plan document, (ii) the employer would have been willing to subsequently modify the payment terms, and (iii) the NQDC plan had given the employee the right to elect to revise the payment schedule.

While this summarizes key NQDC planning opportunities available under the constructive receipt doctrine before Section 409A, two other key sets of income tax rules also impacted the design of NQDC plans: (i) section 83 of the I.R.C. ("Section 83"), and (ii) the economic benefit doctrine.[211]

C. In the Early 1990's, the IRS Encouraged Taxpayers to Establish NQDC Arrangements by Providing Guidance

In many situations, the IRS has disfavored NQDC arrangements, particularly when employees voluntarily defer compensation that they otherwise

---

shareholder . . . is in a position to direct the corporation to do his bidding. . . . In reality, there is nothing standing, [sic] between the taxpayer and the income. It is available at will.").

210. *See ABA Report*, *supra* note 127, at 235-36.

211. Before Section 409A, there were three major hurdles a service provider must clear to avoid immediate taxation on an accrued NQDC benefit: (i) Section 83; (ii) the "constructive receipt" doctrine; and (iii) the "economic benefit" doctrine. The statute and the IRS guidance confirm that these three hurdles continue to apply after the enactment of Section 409A. As a result, even if NQDC is not subject to immediate income taxation under Section 409A, it may be subject to immediate income taxation under any of these three principles. I.R.C. § 409A(c) (2005) ("Nothing in § 409A shall be construed to prevent the inclusion of amounts in gross income under any other provision of this chapter or any other rule of law earlier than the time provided in [§ 409A]."); I.R.S. Notice 2005-1, 2005-2 I.R.B. 274 ("[Section] 409A does not alter or affect the application of any other provision of the Code or common law tax doctrine."). The constructive receipt doctrine is discussed *supra* notes 183-203 and accompanying text. For a discussion of the economic benefit doctrine, see GOLDSTEIN, *supra* note 84, at 48-50. For a discussion of Section 83, see GOLDSTEIN, *supra* note 84, at 46-47.

would receive currently.[212] But in other situations, the IRS has provided taxpayers with a roadmap on how to design a NQDC arrangement that will effectively defer the employee's income tax liability.[213] In 1992, the IRS

---

212. In February 1978, the IRS issued Proposed Treasury Regulation § 1.61-16(a), which provided that if compensation is deferred at the "taxpayer's individual option," the taxpayer will be taxed on the compensation in the tax year in which the option to defer is made, regardless of when the taxpayer actually will receive the compensation. Prop. Treas. Reg. § 1.61-16(a), 43 Fed. Reg. 4638, 4639 (Feb. 3, 1978). Thus, if a 25 year old employee elected to defer an amount to age 65, she would be taxed on the amount immediately, when she is age 25, even though she will not receive the cash for 40 years. Obviously, this would have greatly reduced the appeal of NQDC plans with employee elective deferrals, particularly since the maximum marginal federal income tax rate in 1978 was 70%. I.R.C. § 1(a) (1978) (listing the rate for married individuals filing jointly with taxable income in excess of $203,200). In describing these Proposed Regulations, later commentary stated, "If adopted the 1978 proposed regulations would prohibit employees . . . from participating in unfunded deferred compensation plans as a means of providing tax-deferred retirement income." 1986 BLUE BOOK, *supra* note 101, at 653. In the Revenue Act of 1978, Congress prohibited the IRS from enforcing the Proposed Regulation. Revenue Act of 1978, Pub. L. No. 95-600, § 132, 92 Stat. 2767, 2782 (1978). Basically section 132 of the Revenue Act of 1978 provides that the rules in effect on February 1, 1978, will apply in determining when an employee will be taxed on the benefits under a NQDC plan. The legislative history specifically states, "[T]he doctrine of constructive receipt should not be applied to employees as would be provided in the proposed regulations concerning [NQDC] . . . ." H.R. REP. NO. 95-1445 at 60 (1978), *reprinted in* 1978 U.S.C.C.A.N. 7046, 7097.

213. After a few stinging defeats in court cases, *see supra* note 204 and accompanying text, in the early 1970's the IRS seemed ready to concede certain key points. In *Veit I, Veit II, and Oates*, the IRS argued that once an amount had been earned, it could not be deferred. In General Counsel Memorandum 35196, however the IRS said:

> The position of the Service has generally been that income cannot be deferred once it has been earned. . . .
>
> It is questionable whether the Service could defend such a position in litigation because of the outstanding court cases [citing Veit I, Veit II, and Oates]. . . .
>
> *Based on these cases and the long standing acquiescences, it cannot be stated that a deferral of compensation is only valid if entered into before the performance of services.*

I.R.S. Gen. Couns. Mem. 35,196 (Jan. 16, 1973) (emphasis added).

In addition, in the early 1970's the IRS provided taxpayers with preliminary instructions on how to design a NQDC arrangement. The IRS stated that for a NQDC arrangement to receive an "advance ruling" from the IRS, the plan must meet the following two requirements:

> If the plan provides for an election to defer the payment of compensation, such election must be made before the beginning of the period of service for which the compensation is payable, regardless of the existence in the plan of forfeiture provisions.
>
> If any elections, other than the initial election referred to . . . above, may be made by an employee subsequent to the beginning of the service period the plan must set forth substantial forfeiture provisions that must remain in effect throughout the entire period of the deferral. A substantial forfeiture provision will not be considered to exist unless its conditions impose upon the employee a significant limitation or duty which will require a meaningful effort on the part of the employee to fulfill and there is a definite possibility that the event which will cause the forfeiture could occur.

greatly expanded guidance to taxpayers regarding NQDC payments when it developed the model trust document (often referred to as a "rabbi trust").[214] A "rabbi trust" could be used to provide greater assurance to employees deferring compensation that the NQDC payments would be made as scheduled under the NQDC plan.[215] With such encouragement from the IRS, the number of U.S. employers adopting NQDC plans almost doubled from 1992 to 1996.[216]

Rev. Proc. 71-19, 1971-1 C.B. 698. While Revenue Procedure 71-19 states these two conditions, it also provides that as a general rule "a determination of whether the doctrine of constructive receipt is applicable may be made only after consideration of the specific factual situation involved." *Id.* at 698. Basically, pursuant to such Revenue Procedure, a taxpayer could send a letter to the IRS describing a NQDC arrangement, and obtain a ruling stating whether the arrangement would be effective to defer the employee's income tax liability until the employee actually receives the cash. While only the taxpayer who actually requests and receives the letter ruling is entitled to rely upon it, I.R.C. § 6110(k)(3), the conditions for obtaining a favorable advance letter ruling are frequently used by practitioners in designing other NQDC arrangements.

214. Rev. Proc. 92-64, 1992-2 C.B. 422. A "rabbi trust" is a trust in which the employer contributes deferred compensation which will be invested, and then later distributed to the employee at the time specified in the NQDC plan. It is called a rabbi trust because the first arrangement like this considered by the IRS was established by a congregation for the benefit of its rabbi. *See* Priv. Ltr. Rul. 81-13-107 (Dec. 31, 1980). The IRS described the theoretical foundation for a rabbi trust in I.R.S. Gen. Couns. Mem. 39,230 (May 7, 1984). *See* GOLDSTEIN, *supra* note 84, at 150-51.

Under an IRS model rabbi trust, the employer could transfer the compensation deferred by the employee to a bank or another independent third party (acting as trustee) and the employee would be assured of receiving the NQDC benefits even if (i) the employer subsequently had a "change of heart" and would prefer to pay other creditors before the employee; (ii) there is a "change of control" for the employer (and new management would prefer not to pay the NQDC to the employee); or (iii) the employer has a cash-flow problem. *See* Goldstein et. al., *The Expanding Top Hat: Greater Opportunities with Nonqualified Deferred Compensation*, 55 J. FIN. SERV. PROF'LS 51, 56 (July 1, 2001).

215. In the early 1990's, the IRS was even more generous and provided two exceptions to the requirements set forth in Revenue Procedure 71-19. In addition to allowing an employee to elect to defer before the beginning of the service period, Revenue Procedure 92-65 allows a participant to make an election within thirty days of the establishment of the plan, or within thirty days of the employee becoming eligible to participate in the plan. Rev. Proc. 92-65, § 3.01(a), 1992-2 C.B. 428. Also, the IRS provided four new rules to follow in designing a NQDC arrangement (in order to obtain a favorable IRS ruling):

[1] The plan must define the time and method for payment of deferred compensation for each event (such as termination of employment, regular retirement, disability retirement or death) that entitles a participant to receive benefits. The plan may specify the date of payment or provide that payments will begin within 30 days after the occurrence of a stated event.

[2] The plan may provide for payment of benefits in the case of an "unforeseeable emergency." "Unforeseeable emergency" must be defined in the plan as an unanticipated emergency that is caused by an event beyond the control of the participant or beneficiary and that would result in severe financial hardship to the individual if early withdrawal were not permitted. The plan must further provide that

D. There Were No Extra Penalties for Violating the NQDC Rules Before the Enactment of Section 409A

Prior to the adoption of Section 409A, the IRS did not impose special income tax penalties on employers or employees whose NQDC arrangements violated the constructive receipt doctrine or any other income tax rules.

> *EXAMPLE #3.* In 2003, Gambler, Inc. planned to pay Mr. Lucky a bonus of $10,000. If Mr. Lucky received the bonus, he would pay federal and state income tax, invest the balance, and not spend any of the balance until his retirement in twenty years. Therefore, if Mr. Lucky were in the 40% combined federal and state income tax bracket, he would pay $4,000 in tax immediately and would invest the $6,000 balance for 20 years.
>
> Alternatively, Gambler, Inc. and Mr. Lucky could enter into a NQDC arrangement under which Gambler, Inc. will retain the $10,000, earn interest on it, and pay the balance to Mr. Lucky in 20 years. If the IRS later audits and determines that a Triggering Event in the NQDC plan document caused Mr. Lucky to be taxed on the $10,000 in 2003 under the constructive receipt

---

any early withdrawal approved by the employer is limited to the amount necessary to meet the emergency. Language similar to that described in section 1.457-2(h)(4) and (5) of the Income Tax Regulations may be used.

[3] The plan must provide that participants have the status of general unsecured creditors of the employer and that the plan constitutes a mere promise by the employer to make benefit payments in the future. If the plan refers to a trust, the plan must also provide that any trust created by the employer and any assets held by the trust to assist it in meeting its obligations under the plan will conform to the terms of the model trust, as described in Revenue Procedure 92-64 . . . . Finally, the plan must state that it is the intention of the parties that the arrangements be unfunded for tax purposes and for purposes of Title I of ERISA.

[4] The plan must provide that a participant's rights to benefit payments under the plan are not subject in any manner to anticipation, alienation, sale, transfer, assignment, pledge, encumbrance, attachment, or garnishment by creditors of the participant or the participant's beneficiary.

Id. § 3.01.

216. Drew & Johnston, *supra* note 64, at A1 (indicating that the "[c]umulative number of employers that have notified the Government that they have set up deferred compensation plans" increased from approximately 12,500 in 1992 to approximately 24,000 through Sept. 18, 1996). The article further notes that "[m]any other employers with such plans have not notified the Government" including Exxon, Ford, and Wal-Mart. *Id.* Commentators point out that the enactment of the "$1 Million Cap" of Section 162(m) encouraged many corporations to adopt NQDC plans for their executives earning over $1 million.

> A big push . . . came in 1994, when Congress enacted [Section 162(m) which was] intended to rein in the cost to taxpayers of runaway executive pay. The law barred companies from taking a tax deduction on compensation in excess of $1 million a year for any currently employed individual. So companies encouraged executives to postpone taking the amount of their pay in excess of $1 million until after they left the company or retired.

Schultz & Francis, *supra* note 68, at A9.

> doctrine, Mr. Lucky will be required to pay $4,000 in tax for 2003. Although Mr. Lucky will also pay interest on the $4,000 tax deficiency from the due date of his 2003 tax return (April 15, 2004),[217] as long as the earnings rate on the NQDC balance is at least as high as the IRS interest rate on tax deficiencies, Mr. Lucky will end up in approximately the same economic condition whether he received the compensation immediately in 2003 or deferred the money and violated the constructive receipt doctrine.[218]

Additionally, before the adoption of Section 409A, a faulty attempt to defer could benefit the NQDC plan participants if the applicable statute of limitations expired before the IRS audited and assessed the potential tax liability.[219] But if the statute of limitations period is still open and the IRS successfully argued that the NQDC was taxable to the employee in a prior year, the employer could then argue that it was entitled to an income tax deduction for the NQDC amount in that prior year.[220] Thus, before Section 409A, no great penalty threatened the parties designing a NQDC plan. If the NQDC plan violated the rules, the employee likely would end up in roughly the same economic position as if no attempt had been made to defer the compensation.[221]

IV. ANALYZING SECTION 409A AND ITS PRACTICAL IMPACT—MANDATORY "MAGIC LANGUAGE" AND INFLEXIBILITY

The legislative history of Section 409A provides almost no clues as to the rationale for its enactment. In contrast to other provisions of the 2004 Tax Act, the legislative history of Section 409A contains no "reason for the change" section.[222] Most of the rules flow from the Joint Tax Committee's

---

217. I.R.C. § 6601(a) (requiring that taxpayers pay interest on tax underpayments).

218. This hypothetical scenario operates on the assumption that the typical income tax penalties, which may generally apply in any situation, are inapplicable. *See e.g.*, I.R.C. § 6662 (imposing accuracy-related penalties on underpayments); I.R.C. § 7204 (imposing penalties for fraud).

219. *See, e.g.*, Hornung v. Comm'r, 47 T.C. 428 (1967). In *Hornung*, football star Paul Hornung argued that he constructively received a car he won as an award for being voted the outstanding player of the 1961 National Football League championship game so that the statute of limitations on the IRS' ability to dispute the tax would have expired (the IRS argued that Hornung did not constructively receive the car until 1962). *Id.* at 430-31.

220. *See* I.R.C. § 404(a)(5). An employer is entitled to deduct NQDC when the employee is taxed on the amount, provided that the total amount of the employee's compensation represents "reasonable compensation." *Id.* § 162.

221. Again, this assumes that the NQDC arrangement would not cause the employee to be subject to the fraud penalty, I.R.C. § 7204, the accuracy-related penalty, I.R.C. § 6662, or any other generally applicable income tax penalty.

222. *See* RIA'S COMPLETE ANALYSIS OF THE AMERICAN JOBS CREATION ACT OF 2004, at 2,676, ¶5130 (RIA-Thompson 2004) (discussing in § 811 the "[p]enalty for failing to disclose reportable transactions," relating to I.R.C. § 6707A); *id.* at 2,679, ¶5131 (discussing in § 812 the "[a]ccuracy-related penalty for listed transactions, other reportable transactions having a significant tax-avoidance purpose, etc.," relating to I.R.C. §§ 6662, 6662A, 6664); *id.* at 2,682,

recommendations in the Enron Compensation Report to eliminate flexibility once compensation has been deferred. This Section will consider (A) the disastrous consequences if Section 409A is violated; (B) the broad reach of Section 409A to old and new deferrals; (C) the substantive rules of Section 409A, and (D) the likely impact of Section 409A on the sophisticated and the uninformed.

A. The Consequences of Violating Section 409A—The Triple Tax

Section 409A imposes three significant tax liabilities on the service provider (the employee)[223] if the NQDC plan "is not operated in accordance with [the Section 409A] requirements . . . ."[224] First, "all compensation deferred under the plan for the taxable year and all preceding taxable years shall be includible in gross income for the taxable year [in which the Section 409A violation occurs]."[225] Thus, the employee is subject to income tax immediately on all amounts previously deferred, even if the employer has no obligation to pay the deferred amounts for ten, twenty, thirty, or more years.

Second, if deferrals from prior years are included in taxable income under the first rule, then the income tax for the current year shall be increased by the amount of interest (at the federal underpayment rate[226] plus 1%) "that would

---

¶5132 (discussing in § 813 the "[t]ax shelter exception to confidentiality privileges relating to taxpayer communications," relating to I.R.C. § 7525).

223. Although this Article often refers to the service provider as an "employee," "[t]he application of § 409A is not limited to arrangements between an employer and an employee. For example, § 409A may apply to arrangements between a service recipient and an independent contractor, or arrangements between a partner and a partnership." I.R.S. Notice 2005-1, 2005-2 I.R.B. 277; *see also id.* at 279 (discussing the application of Section 409A to NQDC plans established by a partnership for the benefit of its partners).

224. I.R.C. § 409A(a)(1)(A)(I). The additional amounts that the employee must pay as a result of violating Section 409A are not called "penalties" under the statute; instead, the additional amounts are referred to as additional "taxes." *Id.* § 409A(a)(1)(B).

225. *Id.* § 409A(a)(1)(A)(i)(II). The deferred amounts will not be included in taxable income if the employee's rights to the NQDC benefits are subject to a "substantial risk of forfeiture." *Id.* In connection with Section 83, an amount may be subject to a substantial risk of forfeiture if the employee's right to receive the benefit is contingent on the employee providing "substantial services" for the employer for at least two years. *See* Treas. Reg. § 1.83-3(c), Ex. 1 (2005). A substantial risk of forfeiture typically is *not* included in a NQDC plan when the employee has voluntarily deferred compensation that otherwise would have been paid to her currently. *See supra* note 212. Instead, a substantial risk of forfeiture normally is included only when the employer uses the NQDC plan to provide "supplemental" compensation, or in the case of NQDC plans established by nonprofit employers which are subject to section 457(f) of the I.R.C. ("Section 457(f)"). *See infra* note 260.

226. For example, the federal underpayment rate for the quarter commencing July 1, 2005 was 6% for individuals. Rev. Rul. 2005-35, 2005-24 I.R.B. 1214.

have occurred had the deferred compensation been includible in gross income for the taxable year in which first deferred."[227]

Third, if a deferral from a prior year is included in taxable income under the first rule, then the tax for the current year "shall be increased by . . . an amount equal to 20 percent of the compensation which is required to be included in gross income."[228] This Article will refer to these three additional taxes as the "Triple Tax." The powerful impact of the Triple Tax is demonstrated by the following example:

> *EXAMPLE #4.* Tom Tragic is a middle-manager with taxable income of $180,000 in 2005. For the past 10 years, Tom has voluntarily deferred $10,000 of his annual base salary into the company's NQDC plan, and the company has credited interest at the rate of 5% per annum (compounded annually) on his deferred account balance. Tom and his employer were unaware of the enactment of Section 409A. As a result of minor financial difficulties in 2005, Tom requested that the company distribute $5,000 to him from his NQDC account balance to help pay his daughter's college tuition. The company agreed and made the distribution. Prior to the enactment of Section 409A, Tom would merely be taxed on the $5,000 distribution actually received in 2005 in connection with this transaction. As a result of the enactment of Section 409A, the $5,000 distribution likely will cause Tom to incur an additional federal income tax liability of approximately $89,821 in 2005 (assuming a 7% interest rate discussed below).[229]

---

227. I.R.C. § 409A(a)(1)(B)(ii).

228. *Id.* § 409A(a)(1)(B)(i)(II) (emphasis added).

229. The additional taxable income can be estimated as follows:

(1) Year	(2) Amount Deferred	(3) Additional Interest Earned on NQDC Balance During the Year (sum of col. 4 for all prior years) x 5%*	(4) NQDC Account Increase for the Year	(5) Additional Tax on Increased Deferral (col. (4)) x Income Tax Rate for 2005 (35%)	(6) Interest on Tax Underpayment Equal to the Amount in col. (5) from the Year in col. (1) to 2005**
1995	$ 10,000	-0-	$10,000	$ 3,500	$ 2,389
1996	$ 10,000	$ 500	$10,500	$ 3,675	$ 2,251
1997	$ 10,000	$1,025	$11,025	$ 3,859	$ 2,093
1998	$ 10,000	$1,576	$11,576	$ 4,052	$ 1,914

The extra tax liability arising from a violation of Section 409A can be draconian. In the prior example, Tom received an accelerated distribution of $5,000, and incurred additional tax of $89,821. In this situation, Tom might request additional distributions of a significant portion of his remaining NQDC account balance merely to pay the extra tax triggered by Section 409A, but his employer will have absolutely no obligation to pay him any additional amounts. If the employer does not agree to make additional accelerated distributions (for example, because the employer has its own cash-flow problems), Tom may be forced to take additional measures, such as borrowing on his home, his qualified plan benefits, or perhaps eventually declaring bankruptcy.

Another problem surrounding the Triple Tax is that it applies to "any participant [in the NQDC plan] with respect to whom the failure relates."[230] Thus, if the Section 409A failure is an unpermitted distribution to one employee, the Triple Tax will only apply to the one employee receiving the distribution. But if the problem results from the language in the NQDC plan document, all the participants will be subject to the Triple Tax.[231]

1999	$ 10,000	$2,155	$12,155	$ 4,254	$ 1,712
2000	$ 10,000	$2,763	$12,763	$ 4,467	$ 1,485
2001	$ 10,000	$3,401	$13,401	$ 4,690	$ 1,231
2002	$ 10,000	$4,071	$14,071	$ 4,925	$ 948
2003	$ 10,000	$4,775	$14,775	$ 5,171	$ 633
2004	$ 10,000	$5,513	$15,513	$ 5,430	$ 285
2005	$ 10,000	$6,289	$16,289	$ 5,701	-0-
TOTAL	$110,000	$32,068	$142,068	$49,724	$ 14,941

*The NQDC plan in this example provides that the participant's account balance will earn "interest" at the rate of 5% (compounded annually).

**The "interest" in column (6) of this chart is based on simple interest at an assumed rate of 7%. The actual underpayment rate would be determined under section 6621 of the I.R.C., which provides that the interest rate on tax underpayments is recalculated every calendar quarter, and would be compounded daily as described in Rev. Proc. 95-17, 1995-1 C.B. 556 (superseding Rev. Proc. 83-7, 1983-1 C.B. 583).

The additional tax under Section 409A(a)(1)(A) is represented by column (5), and the additional tax (for the interest) under Section 409A(a)(1)(B)(i)(II), is represented by column (6). Also, the employee will be liable for an additional tax equal to 20% of the sum of the amounts in column (4) before 2005 (20% x $125,779 = $25,156). Thus, the total additional tax will be approximately $89,821 (which is $49,724 + $25,156 + $14,941).

230. I.R.S. Notice 2005-1, 2005-2 I.R.B. 277.

231. *See infra* Part IV.C.1 (regarding some of the precise language which must be used in a

B. *Section 409A Applies Almost All the Time (Except to Tiger Woods)*

The reach of Section 409A is extremely broad.[232] The new rules generally apply to nonqualified amounts deferred after December 31, 2004,[233] and unless the parties proceed with extreme caution, the rules could also apply to amounts deferred before that date. While Section 409A generally applies to NQDC plans established by all employers, including tax-exempt employers,[234] professional golfers may be pleased to hear that the Professional Golfers' Association's NQDC plan is exempt from Section 409A.[235] Since no rationale whatsoever appears for the exception, the PGA's lobbyist apparently scored the political equivalent of a double-eagle![236]

1. Post-2004 Deferrals & Some Exceptions

Section 409A generally applies to "amounts *deferred* after December 31, 2004."[237] An amount is deferred if "the service provider has a *legally binding right* during a taxable year to [the] compensation" and "pursuant to the terms of the plan [the compensation] is payable to . . . the service provider in a later year."[238] Even if an amount is subject to a "substantial risk of forfeiture,"[239] the

---

NQDC plan to avoid Section 409A).

232. The legislative history provides, "A nonqualified deferred compensation plan is any plan that provides for the deferral of compensation other than a qualified employer plan or any bona fide vacation leave, sick leave, compensatory time, disability pay, or death benefit plan." Conference Agreement, *supra* note 13, at 734.

233. American Jobs Creation Act of 2004, Pub. L. No. 108-357, § 885(d), 118 Stat. 1418 (2004).

234. I.R.S. Notice 2005-1, 2005-2 I.R.B. 279.

235. *See* American Jobs Creation Act of 2004 § 885(d)(3),

236. The statutory language of the exception indicates its arbitrariness. The exception is available for any nonelective NQDC plan established or maintained by a tax-exempt organization which was incorporated on July 2, 1974, if the plan was in existence on May 1, 2004. *See* American Jobs Creation Act of 2004 § 885(d)(3).

237. *Id.* § 885(d)(1) (emphasis added).

238. I.R.S. Notice 2005-1, 2005-2 I.R.B. 277 (emphasis added).

239. The Triple Tax generally cannot apply as long as the NQDC is "subject to a substantial risk of forfeiture." I.R.C. § 409A(a)(1)(A)(i)(II) (2005). But Congress has given the Secretary of the Treasury authority to issue regulations "disregarding a substantial risk of forfeiture in cases where necessary to carry out the purposes of [Section 409A]." I.R.C. § 409A(e)(5). Preliminary guidance from the Treasury indicates that a "substantial risk of forfeiture" for purposes of Section 457(f) may not be a "substantial risk of forfeiture" under Section 409A. For example, the IRS states,

> [A]ny extension of a period during which compensation is subject to a substantial risk of forfeiture . . . whether elected by the service provider, service recipient or other person *(or by agreement of two or more of such persons)* is disregarded for purposes of determining whether such compensation is subject to a substantial risk of forfeiture. . . . For purposes of §409A, an amount will not be considered subject to a substantial risk of forfeiture beyond the date or time at which the recipient otherwise could have elected to receive the

service provider is still considered to have a "legally binding right"[240] to the amount. Section 409A can apply to closely-held corporations, public corporations, and as the legislative history expressly provides, tax-exempt organizations.[241]

Because Section 409A is designed to apply to "nonqualified" deferred compensation plans, "qualified" plans, such as profit sharing plans and pension plans described in Section 401(a), are excluded.[242] Section 409A can apply to "equity-based" NQDC arrangements, including stock options, stock appreciation rights, and stock purchase plans,[243] but exceptions are provided for incentive stock options described in section 422 of the I.R.C. and nonqualified stock options that are not "in the money" when issued.[244]

---

amount of compensation . . . . For example, a salary deferral generally may not be made subject to a substantial risk of forfeiture.

I.R.S. Notice 2005-1, 2005-2 I.R.B. 280 (emphasis added).

240. *Id.* at 277. Apparently the "legally binding right" language was included to exclude "compensation [which] may be unilaterally reduced or eliminated by the service recipient or other person after the services creating the right to the compensation have been performed." *Id.* But "compensation is not considered subject to unilateral reduction or elimination merely because it may be reduced or eliminated by operation of the objective terms of the plan, such as the application of an objective provision creating a substantial risk of forfeiture . . . ." *Id.*

241. Conference Agreement, *supra* note 13, at 735. The legislative history excludes "eligible" plans of tax-exempt organizations governed by section 457(b) of the I.R.C., but provides that "[a] tax-exempt or governmental deferred compensation plan that is not an eligible deferred compensation plan is not a qualified employer plan," which would not be subject to Section 409A. *Id.* As a result, tax-exempt or governmental NQDC plans governed by Section 457(f) can be subject to Section 409A. *See infra* notes 260-61 and accompanying text (discussing when a tax-exempt employer would use a Section 457(f) plan); *see also* GOLDSTEIN, *supra* note 84, at 190-94 (discussing ways to create a Section 457(f) plan that contains a substantial risk of forfeiture).

242. I.R.C. § 409A(d)(1). Eligible deferred compensation plans for governmental and tax-exempt employers described in section 457(b) of the I.R.C. are treated as "qualified plans," and therefore are exempt from Section 409A. I.R.C. § 409A(d)(2)(B); *see also* Conference Agreement, *supra* note 13, at 734-35 (emphasizing the intent to exclude qualified plans from this legislation). Other exceptions from Section 409A include "any bona fide vacation leave, sick leave, compensatory time, disability pay, or death benefit plan." I.R.C. § 409A(d)(1)(B); *see also* I.R.S. Notice 2005-1, 2005-2 I.R.B. 277 (emphasizing that these exceptions must be *"bona fide"*). Additionally, an exception for short-term deferrals excludes bonuses paid within 2-1/2 months of the end of the year, and amounts paid within 2-1/2 months of when an amount is no longer subject to a "substantial risk of forfeiture" for purposes of Section 409A. *Id.* at 277-78.

243. I.R.S. Notice 2005-1, 2005-2 I.R.B. 278.

244. For purposes of this Article, the term "in the money" is used to describe a stock option that has been granted to the employee if the fair market value of the stock at the time exceeds the price that the employee must pay in order to purchase the stock (the "exercise price"). In excluding qualified stock options (also called "incentive stock options") the legislative history states, "[409A] is not intended to change the tax treatment of incentive stock options meeting the requirements of 422 or options granted under an employee stock purchase plan meeting the

2. When Pre-2005 Deferrals Can Become Subject to Section 409A—the "Material Modification" Rules

While Section 409A typically applies to amounts deferred after 2004,[245] pre-2005 deferrals are subject to Section 409A "if the plan under which the deferral is made is *materially modified* after October 3, 2004," unless an exception applies.[246] While arrangements often lose their "grandfathered" status under the tax laws as a result of a material modification,[247] an alarming feature of Section 409A is that the parties can unintentionally trigger a material modification very easily. The legislative history provides that "[t]he addition of any benefit, right or feature is a material modification."[248] As a result, the mere deferral of additional amounts under the terms of the existing NQDC plan will be a material modification that triggers the application of Section 409A unless an exception applies.[249]

Consequently, parties have already inadvertently and possibly unknowingly materially modified the NQDC plan if they continued to defer compensation

---

requirements of section 423." Conference Agreement, *supra* note 13, at 735. In excluding stock options that are not in the money, the legislative history states,

> [I]t is not intended that the term "nonqualified deferred compensation plan" include an arrangement taxable under section 83 providing for the grant of an option on employer stock with an exercise price that is not less than the fair market value of the underlying stock on the date of grant if such arrangement does not include a deferral feature other than the feature that the option holder has the right to exercise the option in the future.

Conference Agreement, *supra* note 13, at 735; *see also* I.R.S. Notice 2005-1, 2005-2 I.R.B. at 278 (discussing nonstatutory and statutory stock options).

245. American Jobs Creation Act of 2004, Pub. L. No. 108-357, § 885(d)(1), 118 Stat. 1418 (2004).

246. *Id.* § 885(d)(2) (emphasis added).

247. *See, e.g.*, I.R.S. Field Serv. Adv., 2002 WL 1315773 ("If a liability is materially modified, the liability loses its grandfathering and becomes subject to a different set of regulations, depending on when the material modification occurs."); I.R.S. Tech. Adv. Mem. 1999-03-032 (Oct. 2, 1998) ("[A] plan's grandfather status ends, and [I.R.C.] section 457 will apply, as of the first effective date of the first material modification of the plan agreed to after December 31, 1987."); Disallowance of Deductions for Employee Remuneration in Excess of $1,000,000, 58 Fed. Reg. 66,310 (Dec. 20, 1993) (notice of proposed rulemaking) ("If a contract is materially modified, then the contract is no longer grandfathered and any amounts that have not yet been received under the contract are subject to section 162(m).").

248. Conference Agreement, *supra* note 13, at 737. The legislative history provides examples of a material modification, including an acceleration of a vesting provision, and a "haircut" provision that requires employees to forfeit 10% of the requested amount. *Id.* But the history states, "A change in the plan administrator would not be a material modification." *Id.*

249. I.R.S. Notice 2005-1, 2005-2 I.R.B 285 ("[T]he grant of an additional benefit under an existing arrangement that consists solely of a deferral of additional compensation not otherwise provided under the plan as of October 3, 2004 will be treated as a material modification of the plan only as to the additional deferral of compensation, *if* the plan explicitly identifies the additional deferral of compensation and provides that the additional deferral of compensation is subject to § 409A." (emphasis added)).

after October 3, 2004. Section 409A applies to all amounts deferred unless parties amended their plans to explicitly identify the compensation deferred after October 3, 2004 and provide that those deferrals are subject to Section 409A.[250] If additional amounts are deferred into a pre-2005 NQDC plan, the parties *must* amend the plan to avoid the Triple Tax on all amounts deferred. Section 409A demands that the NQDC plan amendments contain precise wording (referred to in this Article as "magic language") in order to avoid the Triple Tax.

> *EXAMPLE #5.* As in Example #4 above, Tom Tragic is a participant in his employer's NQDC plan. Under the plan, Tom defers $10,000 of his salary each year for eleven years (from 1995 through 2005). Assume that the NQDC plan was prepared several years ago, and does not include the "magic language" required by Section 409A.[251] As of October 3, 2004, Tom's account balance was $125,779. Tom and his employer continue to defer and they do not amend the plan by December 31, 2005 (or another date specified by the IRS) to include such magic language. In this case, the NQDC plan has been materially modified because the deferrals after October 3, 2004 entitle Tom to additional benefits and the plan was not amended on or before the deadline to comply with Section 409A.[252] Since the NQDC plan does not comply with Section 409A, Tom will be subject to the Triple Tax, which may be approximately $89,821, as shown in Example #4 above.[253] Presumably, Tom will be subject to the Triple Tax in 2005.[254]

As another example of the broad reach of Section 409A with respect to old deferrals, any amount deferred before 2005 that was not *vested* by December 31, 2004, will be subject to Section 409A.[255] Any amount subject to a

---

250. I.R.S. Notice 2005-1, 2005-2 I.R.B. 284. This affirmative duty to act is reinforced by the IRS' statement:

> A plan adopted before December 31, 2005, will not be treated as violating § 409A(a)(2), (3), or (4) *only if* (i) the plan is operated in good faith compliance with the provisions of § 409A and this notice during the calendar year 2005, and (ii) *the plan is amended on or before December 31, 2005 to conform to the provisions of § 409A with respect to amounts subject to § 409A.*

Id. at 285 (emphasis added).

251. Section 409A requires that certain terms used in NQDC plans, such as "disability," "separation from service," and "change of control," must conform to definitions prescribed by the statute or IRS guidance. *See infra* Part IV.C.1.

252. *See supra* note 249 and accompanying text.

253. *See supra* note 229 and accompanying text.

254. Because Section 409A generally applies to "amounts deferred after December 31, 2004," the Triple Tax presumably cannot apply in 2004, even if amounts are deferred after October 3, 2004. American Jobs Creation Act of 2004, Pub. L. No. 108-357, § 885(d)(1), 118 Stat. 1418 (2004).

255. Conference Agreement, *supra* note 13, at 737 ("For purposes of the effective date, an amount is considered deferred before January 1, 2005, if the amount is earned and *vested* before such date." (emphasis added)); *see also* Prop. Treas. Reg. § 1.409A-6(a)(1), 70 Fed. Reg.

substantial risk of forfeiture will not be considered vested for these purposes. As previously discussed, an employer will frequently make supplemental deferred compensation subject to a substantial risk of forfeiture to impose golden handcuffs on the employee.

> *EXAMPLE #6.* Ellen Executive is very valuable to the company. From 1994 through 2003, the company added $10,000 annually to a NQDC account maintained for Ellen. These amounts represent an addition to Ellen's regular salary and other benefits—Ellen did not voluntarily defer a portion of the compensation she otherwise would have received. With earnings, her account balance had grown to $125,000 by December 31, 2003. Under the company-designed NQDC plan, Ellen will forfeit the NQDC benefit if she quits before attaining age 65, or if she is fired for cause.[256] But if she remains a full-time employee until age 65, she will receive installment payments over a period of 10 years. Ellen will turn 65 in 2007.

This type of economic arrangement provides a greater incentive for Ellen to stay with the company as she grows older—a classic golden handcuff arrangement.[257] Although the amounts were credited to her NQDC account balance before December 31, 2004, these amounts would not be considered vested before the effective date of Section 409A, due to the substantial risk of forfeiture.[258] In this example, Ellen's account balance will not vest until 2007; as a result of the substantial risk of forfeiture, the plan will be subject to Section 409A. Thus, the NQDC plan must be amended to avoid the Triple Tax of Section 409A.[259]

The NQDC benefits for employees of tax-exempt organizations must be subject to a substantial risk of forfeiture when the amount deferred during the year exceeds a certain amount (generally $14,000 in 2005).[260] Consequently, such plans will all be subject to Section 409A and will not be grandfathered in even if no amounts are deferred after 2004.[261] This can be a trap for tax-exempt organizations, their employees, and other service providers.

57930, 57982 (Oct. 4, 2005); I.R.S. Notice 2005-1, 2005-2 I.R.B. 275 ("[A]n amount will be treated as deferred on or before December 31, 2004 only if the service recipient has a binding legal obligation to pay an amount in a future taxable year *and* the service provider's right to the amount is earned *and vested* as of December 31, 2004." (emphasis added)).

256. If the plan contains a provision that the employee will forfeit the benefit if she quits or is fired for cause, the employee benefit will be subject to a substantial risk of forfeiture, even though the employee may receive the benefit if she is fired without cause. Priv. Ltr. Rul. 99-43-008 (Oct. 29, 1999).

257. *See supra* note 181 and accompany text.

258. *See supra* note 255.

259. In this scenario, Ellen's plan contains no magic language.

260. *See* I.R.C. § 457 (2005). When the amount deferred exceeds the applicable dollar amount, the NQDC benefits must be subject to a substantial risk of forfeiture for the participants to avoid an immediate tax liability on the deferred amounts. *Id.*

261. Since the NQDC amounts will be subject to a substantial risk of forfeiture, the

The only NQDC plans that are not subject to Section 409A are plans for which (i) no deferrals are made after October 3, 2004, (ii) all deferred amounts were vested, and (iii) no additional benefit, right or feature is added after October 3, 2004. Plans that meet these three criteria will continue to be subject to the constructive receipt doctrine and any other rules applicable to NQDC other than Section 409A.[262] To avoid the Triple Tax, all other pre-2005 NQDC plans must be amended in a timely manner to include the magic language mandated by Section 409A.

3. Earnings Are Subject to Section 409A if the Principal Is Subject to Section 409A

The earnings accrued on deferred compensation can add up to significant amounts. At an interest rate of 7%, an account balance will double approximately every ten years. The IRS has stated that if employers were entitled to claim an income tax deduction on interest accrued on NQDC accounts that are not paid to employees each year, the government would lose approximately $7 billion in lost tax revenue.[263] NQDC arrangements typically provide that the employee's deferred compensation account balance will accrue interest until the entire deferred compensation account balance is paid to the employee.[264] Some companies apply an excessively generous earning rate for

---

amounts will not be considered vested for purposes of Section 409A's effective date rule until the substantial risk of forfeiture lapses. *See supra* note 255.

262. The legislative history provides,

> Operating under the terms of a deferred compensation arrangement that complies with current law and is not materially modified after October 3, 2004, with respect to amounts deferred before January 1, 2005, is permissible, as such amounts would not be subject to the requirements of [Section 409A]. For example, subsequent deferrals with respect to amounts deferred before January 1, 2005, under a plan that is not materially modified after October 3, 2004, would be subject to present law, and would not be subject to [Section 409A].

Conference Agreement, *supra* note 13, at 737.

263. *See* Albertson's Inc. v. Comm'r 42 F.3d 537, 539 (9th Cir. 1994).

264. The parties have flexibility in determining the amount of earnings that will accrue on the employee's NQDC account balance. The NQDC document may fix an earnings rate that will apply throughout the term of the arrangement, it may tie the earnings rate to a fluctuating interest rate (such as the prime rate), or the employer may retain discretion to set the interest rate each year.

A frequently used technique that has been sanctioned by the IRS permits the employee to select investments, which may include blue-chip stocks, bonds, mutual funds, indexes, and other investments, that can be valued easily from a list established by the employer. Priv. Ltr. Rul. 95-05-012 (Nov. 4, 1994). "[T]he participants' right to designate investments in which their deferrals are deemed to be invested under the Plans will [not] cause these amounts to be includible in the gross income of the participants until the taxable year in which those amounts are actually paid or made available to them." *Id.* The employee's NQDC account balance will increase (or decrease) *as if* the employee's NQDC account balance had been invested solely

top executives. For example, the New York Times once reported that General Electric credited CEO Jack Welch's NQDC account balance with a 14% annual rate of interest.[265]

---

based upon the employee's designations. *Id.* The employer, however, is not *obligated* to invest the amounts deferred according to the employee's designations. Instead, the employer can invest in a different manner and attempt to select investments that may perform better than the employee's investment choices.

EXAMPLE #7. Gambler, Inc. establishes a NQDC plan under which its employees can voluntarily elect to defer a portion of their base salaries and annual bonuses, and designate investments from a list of ten different publicly-traded mutual funds. Mr. Lucky, an employee of Gambler, Inc., elects to defer $10,000 in Year 1 and designates that all of his deferrals shall grow (or decline) based on the investment performance of one of the ten mutual funds, the High Flyer Fund. Mr. Lucky defers no other amounts over a period of ten years and the High Flyer Fund grows at an average rate of 20% per year compounded annually. Mr. Lucky retires at the end of the ten-year period. As a result of the performance of the High Flyer Fund, Mr. Lucky will be entitled to $61,917. If Gambler, Inc. invested the $10,000 originally deferred in the High Flyer Fund, Gambler, Inc. will have $61,917, which it can pay to Mr. Lucky under the NQDC plan. In contrast, if Gambler, Inc. did not follow Mr. Lucky's investment designations, the employer may have either a wind-fall or a short-fall. If the investments chosen by Gambler, Inc. earned 10%, Gambler, Inc. would have only $25,937 from which to pay the $61,917 NQDC benefit to Mr. Lucky. On the other hand, if investments chosen by Gambler, Inc. earned 30% per year, Gambler, Inc. would have $137,858 from which to pay the NQDC benefit of $61,917 to Mr. Lucky, and Gambler, Inc. could retain the excess of $75,941.

Note that if Gambler, Inc. follows Mr. Lucky's investment designations, and it sells the High Flyer Fund interests to pay Mr. Lucky, or distributes the High Flyer Fund interests in-kind to Mr. Lucky, Gambler, Inc. will be obligated to pay income tax on the gain. I.R.C. § 1001(a) (2005).

NQDC arrangements can be structured as either account balance plans or defined benefit plans. When an employee voluntarily defers her own compensation, the arrangement usually will be an "account balance plan"—the employer will establish a hypothetical bookkeeping account, which will equal the total amounts deferred, plus any deemed investment earnings, minus any deemed investment losses, minus any benefit payments. Thus, the participant will have a NQDC account balance on the employer's books. In contrast, an employer providing *supplemental* deferred compensation may use a defined benefit approach. For example, if the employee remains a full-time employee until she turns 65, she will receive a NQDC benefit of $5,000 per month for ten years. Here the employee has no account balance, but is entitled to a defined benefit if certain conditions are satisfied.

A NQDC plan is an account balance plan if it meets three requirements: "[1] a principal amount . . . is credited to an individual account for an employee, [2] the income attributable to each principal amount is credited . . . to the individual account, and [3] the benefits payable to the employee are based solely on the balance credited to the individual account." Treas. Reg. § 31.3121(v)(2)-1(c)(1)(ii) (2005). Any NQDC plan that does not meet these requirements is a nonaccount balance plan. *Id.* § 31.3121(v)(2)-1(c)(2)(i). An example of a nonaccount balance plan is a defined benefit plan providing that the employee will receive a retirement benefit equal to the product of (i) 2% for each year of service, and (ii) the employee's highest average annual compensation for a three-year period. *Id.* § 31.3121(v)(2)-1(c)(4), Ex. (5)(i).

265. Drew & Johnston, *supra* note 64, at A1. In addition, "Many deferral plans pay 8 to 12

Under Section 409A, "[r]eferences to the deferral of compensation include references to income (whether actual or notational) attributable to such compensation or such income."[266] To avoid the Triple Tax, both the principal amounts deferred, and the earnings on those amounts, must satisfy the requirements of Section 409A.

Regarding the effective date rules, Section 409A "shall apply to earnings on deferred compensation only to the extent that [Section 409A] appl[ies] to such compensation."[267] Arguably, if the principal amounts deferred before 2005 are subject to Section 409A because those amounts were not vested before 2005,[268] the earnings on those amounts likewise would be subject to Section 409A, even if the earnings were vested.[269]

---

percent interest and some go higher. BellSouth has paid interest rates from 8 to 22 percent on deferrals . . . ." *Id.* Pharmaceutical giant Wyeth guaranteed its executives a 10% return on their deferred pay. Schultz & Francis, *supra* note 68, at A1 ("Wyeth paid [its Chairman John R.] Stafford $3.8 million in interest on a deferred-compensation account valued at $37.75 million . . . ."). Boise Cascade credits its executives 8.9% in interest; Harrah's Entertainment provides a 12% to 13% interest rate on its NQDC plans; and in 2001, Halliburton Co. paid an average of 9.8% interest on deferrals. *Id.* at A9. "One of Enron's plans guaranteed executives a minimum 12% return on deferrals." *Id.* Special FICA tax rules apply when a fixed interest rate in excess of Moody's Average Corporate Bond Yield is used. Treas. Reg. § 31.3121(v)(2)-1(d)(3), Ex. 4 (2005).

266. I.R.S. Notice 2005-1, 2005-2 I.R.B. 279; *see also* I.R.C. § 409A(d)(5) (setting forth the rule for treatment of earnings). Presumably the term "notational" refers to arrangements similar to the one considered in Private Letter Ruling 95-05-012 in which the employee's balance grows based on performance of the employee's hypothetical investment designations, but the employer has no obligation to actually invest amounts according to the employee's designations. *See supra* note 261.

267. American Jobs Creation Act of 2004, Pub. L. No. 108-357, § 885(d)(2)(A), 118 Stat. 1418 (2004).

268. *See supra* note 255 and accompanying text (discussing when amounts are considered vested for purposes of the effective date rule).

269. This scenario likely will apply to Section 457(f) NQDC plans established by tax-exempt employers. The principal amounts deferred under a Section 457(f) plan must be subject to a substantial risk of forfeiture in order for the employee to escape immediate taxation on the amounts deferred. But under the applicable Treasury regulations, the earnings on the amounts deferred could be immediately vested and yet the employee would not be subject to income tax on the earnings until the earnings are actually paid, or until there is no longer a substantial risk of forfeiture. *See* Treas. Reg. § 1.457-11(a)(2)-(3) (2005). The wording of Section 409A(d)(2) may support the argument that whenever the principal amounts deferred are subject to Section 409A, the earnings also will be subject to Section 409A, even if the employee's rights to the earnings were vested.

C. *The Substantive Rules of Section 409A—Timing Rules that Eliminate Flexibility*

In response to the Enron Compensation Report's urge to end flexibility with respect to the timing of NQDC payments,[270] Section 409A includes a series of rules that mandate the circumstances under which payments can be made, prohibit accelerating payments, and restrict the ability to subsequently delay payments.

1. The Six Permissible Payment Events and the Imposition of "Magic Language" Requirements

Under Section 409A, the Triple Tax will apply if payments under a NQDC plan are to be made at any time, unless such payments are contingent upon one of the following six "Payment Triggers":

1. disability;
2. separation from service;
3. death;
4. "a specified time (or pursuant to a fixed schedule) specified under the plan at the date of the deferral of such compensation;"[271]
5. a change of control of the employer entity; or
6. the occurrence of an unforeseeable emergency.[272]

Under Section 409A, plans may allow for payments to be made upon the first occurrence of any, or all, of these events.[273] For example, a NQDC plan might provide for payments when an employee turns 65 or separates from service, whichever occurs first, but the plan might also provide that no distributions would be made upon a change of control or in the event of an unforeseeable emergency. Thus, while distributions can only be made upon the occurrence of one of the six specified Payment Triggers, an employer can pick and choose between the six Payment Triggers in designing a NQDC plan.

Section 409A's rigid allowance of only six Payment Triggers presents a major limitation on plan design. In this respect, Section 409A is in sharp contrast with IRS Revenue Ruling 60-31, other IRS rulings, and several court

---

270. *See supra* notes 143-45 and accompanying text.

271. I.R.C. § 409A(a)(2)(A)(iv) (2005).

272. The six permissible Payment Triggers are listed in Section 409A(a)(2)(A). The exclusivity of the six permissible Payments Triggers is emphasized in the legislative history: "A nonqualified deferred compensation plan may not allow distributions other than upon the permissible distribution events . . . ." Conference Agreement, *supra* note 13, at 730.

273. Conference Agreement, *supra* note 13, at 733. "It is intended that multiple payout events are permissible. For example, a participant could elect to receive 25 percent of their account balance at age 50 and the remaining 75 percent at age 60." *Id.*

cases, which previously permitted NQDC plans to include many different types of payment triggers, including (i) the employee becoming partially incapacitated; (ii) a full-time employee becoming a part-time employee; or (iii) the employer's net worth falling below $10 million.[274]

The level of inflexibility imposed by Section 409A is epitomized by four of the six Payment Triggers, discussed in subsections (a) through (d) below. Each of these Payment Triggers require precise "magic language." If a drafter of a NQDC plan fails to include the magic language in the plan document, the plan participants may be subject to the Triple Tax.

a. The Ridiculous Disability Rule

First, Section 409A mandates a precise definition of "disability," which will have harsh consequences. Genuinely disabled employees whose unique situations do not fit this narrow definition may not be considered "disabled" for Section 409A purposes.

> [A] participant shall be considered disabled if the participant—
> (i) is unable to engage in any *substantial gainful activity* by reason of any medically determinable physical or mental impairment which can be expected to result in death or can be expected to last for a continuous period of not less than 12 months, or
> (ii) is, by reason of any medically determinable physical or mental impairment which can be expected to result in death or can be expected to last for a continuous period of not less than 12 months, receiving income replacement benefits for a period of not less than *3 months* under an accident and health plan covering employees of the participant's employer.[275]

Part (i), requiring an inability to engage in substantial gainful activity, may be a very difficult standard to satisfy.[276] Thus, as a practical matter, the second

---

274. *See supra* notes 189-201 and accompanying text.

275. I.R.C. § 409A(a)(2)(C) (emphasis added).

276. While Section 409A does not define "unable to engage in any substantial gainful activity," the phrase is also used as a standard for determining whether an individual is eligible to receive social security disability payments. *See* 42 U.S.C. § 423(d)(1)(A) (2000) (defining disability in part, as the "inability to engage in any substantial gainful activity by reason of any medically determinable physical or mental impairment which can be expected to result in death or which has lasted or can be expected to last for a continuous period of not less than 12 months"); *see also* 20 C.F.R. § 404.1505(a) (2005) ("The law defines disability as the inability to do any substantial gainful activity by reason of any medically determinable physical or mental impairment which can be expected to result in death or which has lasted or can be expected to last for a continuous period of not less than 12 months."). A presumption arises that an individual can "engage in substantial gainful activity" if she earns $830 or more per month (in 2005). Update 2006, SSA Publication No. 05-10003 (Jan. 2006). Thus, an individual could be deemed to be engaged in substantial gainful activity even if she can merely perform a couple of hours work at a minimum wage job each day. As a result, a former top executive who can now

part of the definition, which may come to be known as the "Three Month Rule," likely will be the operative standard. Three months may seem like a reasonable time frame to use when determining whether an employee is seriously disabled. But the adoption of a bright-line test, coupled with the Triple Tax penalty for violating Section 409A, imposes a harsh sanction if a pre-existing NQDC arrangement is not amended before the IRS deadline.

> *EXAMPLE #8.* Same facts as in Example #4, except this time Tom does not request a distribution. The NQDC plan document provides that Tom is entitled to begin receiving distributions once he is disabled, which begins when he becomes entitled to payments under the company's long-term disability plan maintained for employees generally. Tom defers additional amounts under the plan in 2008, and the plan document is not amended to comply with the Three Month Rule of Section 409A.[277] As a result, the terms of the NQDC plan do not satisfy Section 409A, and Tom will be subject to the Triple Tax of Section 409A(a)(1) in 2008 on all amounts ever deferred under the NQDC plan.[278]

The Three Month Rule also creates a significant risk when a simple employment agreement providing for NQDC is drafted without regard to Section 409A.

> *EXAMPLE #9.* Larry Labor Attorney drafts an employment agreement between Tom Tragic and his employer in 2008. The agreement provides that $10,000 of Tom Tragic's salary will be deferred and paid to him upon the earlier of his death, disability, or separation from service. Under the agreement, Tom is considered disabled when he *becomes* eligible to receive payments under the employer's long-term disability plan. Tom will be subject to the Triple Tax of Section 409A(a)(1) in 2008 on the deferred amount because the plan document does not satisfy the Three Month Rule even though he may not receive the deferred amount for many years.[279] The Triple Tax could have been avoided had Larry Labor Attorney drafted the employment agreement to incorporate the Section 409A(a)(2)(C) definition of "disability."

As frequently occurs with bright-line tests, the results can be bizarre in situations that are close to the line. In the example above, if Larry Labor Attorney had defined disability as receiving benefit payments for two months instead of three months, the Triple Tax still would apply.

---

only make change at a restaurant for a few hours a day may not be considered disabled under this substantial gainful activity standard.

277. *See generally* I.R.S. Notice 2005-1, 2005-2 I.R.B. 274 (providing guidance for holders of NQDC plans in light of Section 409A).

278. *See supra* note 229 and accompanying text (describing the impact of the Triple Tax).

279. In this case, the Triple Tax will cost Tom $89,821. *See supra* note 229 and accompanying text.

b. Separation from Service

As with the term disability, a specialized definition of "separation from service" will be required to satisfy Section 409A(a)(2) and avoid the Triple Tax. Congress did not provide a definition of separation from service, but instead directed the Secretary of the Treasury to provide the definition.[280] Potential areas of concern could include (i) employees who terminate employment with the sponsor of the NQDC plan, but return to work for an affiliated entity, and (ii) employees who provide minimal consulting services after terminating full-time employment.[281]

> *EXAMPLE #10.* Old Reliable has worked for the company for over forty years in many capacities and defers compensation under the company's NQDC plan. Old Reliable desires to retire from full-time employment, but would like to work at the office a few mornings each week. The company would be delighted if Old Reliable would be available, even occasionally, because of her vast experience. The rules of Section 409A can, however, interfere with the economic decision that Old Reliable will make. [282] If the NQDC Plan is drafted in compliance with Section 409A and Old Reliable wishes to begin receiving NQDC payments (and avoid the Triple Tax), she will be required to terminate employment completely, even though both she and her employer wish for her to continue working part-time.[283] If Old Reliable continued to work part-time and the company began making NQDC payments to Old Reliable, the plan would not be *operated* in accordance with Section 409A, and Old Reliable would be subject to the Triple Tax.[284]

In connection with the description of a separation from service, Congress imposed a special timing rule with regard to public companies. Basically, in the case of a "key employee"[285] of a public corporation, payments triggered by

---

280. I.R.C. § 409A(a)(2)(A)(I).

281. The legislative history provides that "it is intended that aggregation rules would apply in the case of separation from service so that the separation from service from one entity within a controlled group, but continued service for another entity within the group, would not be a permissible distribution event." Conference Agreement, *supra* note 13, at 736.

282. A NQDC benefit that was earned and vested prior to January 1, 2005, could be "grandfathered" under the rules in effect prior to the enactment of Section 409A. American Jobs Creation Act of 2004, Pub. L. No. 108-357, § 885(d), 118 Stat. 1418 (2004). Under the pre-Section 409A rules, Old Reliable could begin receiving NQDC payments when she switches from full-time to part-time employment. Rev. Rul. 60-31, 1960-1 C.B. 175.

283. In this example, the payment of benefits to Old Reliable would violate the terms of the NQDC plan.

284. I.R.C. § 409A(a)(1)(A)(i)(II) (providing that a NQDC plan that is not "operated in accordance with such requirements" violates Section 409A).

285. The legislative history states that "[k]ey employees . . . generally include officers having annual compensation greater than $130,000 (adjusted for inflation and limited to 50 employees), five percent owners, and one percent owners having annual compensation from the employer greater than $150,000." Conference Agreement, *supra* note 13, at 730 n.811.

a separation from service cannot begin until at least six months after the separation from service.[286]

c. Change in Control

Congress also directed the Secretary of the Treasury to provide a definition to specify when a "change in control" of the employer has occurred that will permit a distribution from a NQDC plan.[287] Three types of events can constitute a "change of control": (i) a "change in the ownership of a corporation," (ii) a "change in the effective control of the corporation," or (iii) a "change in the ownership of a substantial portion of the assets of the corporation."[288] The Treasury Department's initial guidance uses a series of bright-line tests to define a change of control. These tests basically include: (i) the acquisition of 50% of the fair market value (or voting power) of the corporation's stock; (ii) the involuntary replacement of 35% or more of the directors, or (iii) the acquisition of 40% or more of the total assets.[289]

Establishing bright-line tests for when a NQDC payment is permitted creates the risk that uninformed taxpayers will inadvertently trigger the Triple Tax. The Triple Tax will be triggered if a NQDC plan subject to Section 409A provides that benefits will be paid, for example, (i) on a transfer of 40% of the fair market value of the corporation's stock; (ii) on a transfer of 40% of the corporation's voting stock; (iii) the unendorsed replacement of 30% of the

---

286. I.R.C. § 409A(a)(2)(B)(i). Presumably this rule was designed to further restrict the flexibility available in making payments.

287. *Id.* § 409A(a)(2)(A)(v).

288. I.R.S. Notice 2005-1, 2005-2 I.R.B. 280; *see also* Prop. Treas. Reg. § 1.409A-3(g)(5), 70 Fed. Reg. 57930, 57977 (Oct. 4, 2005) (setting forth proposed NQDC rules for changes in ownership, effective control, or ownership of substantial assets of a corporation).

289. I.R.S. Notice 2005-1, 2005-2 I.R.B. 281. The Treasury Department Notice provides the following guidance regarding a change in control of the ownership of a corporation:

> [A] change in the ownership of a corporation occurs on the date that any one person, or more than one person acting as a group . . . acquires ownership of stock of the corporation that, together with stock held by such person or group, constitutes more than 50 percent of the total fair market value or total voting power of the stock of such corporation.

Id.

Second, a "change in the effective control of a corporation" occurs if (i) any one person or group acquires, within a 12-month period, 35% or more of the total voting power of stock of the corporation, or (ii) "a majority of members of the corporation's board of directors is replaced during any 12-month period by directors whose appointment or election is not endorsed by a majority of the members of the corporation's board of directors prior to the date of appointment or election." *Id.*

Third, a "change in the ownership of a substantial portion of a corporation's assets" occurs when any one person (or a group) acquires, within a 12-month period, "assets from the corporation that have a total gross fair market value equal to or more than 40 percent of the total gross fair market value of all of the assets of the corporation immediately prior to such acquisition or acquisitions." *Id.* at 282.

corporation's directors, or (iv) a transfer of 35% of the assets.[290] Thus, the mere failure of the plan scrivener to follow the precise linguistics of Section 409A can have disastrous consequences for the plan participants (namely, the Triple Tax).

d. The Specified Time Rule

NQDC distributions at "a specified time (or pursuant to a fixed schedule) specified under the [NQDC] plan at the date of the deferral of such compensation" are permitted without triggering the Triple Tax.[291] Contrary to established case law before the enactment of Section 409A,[292] however, distributions cannot be triggered by the occurrence of an *event,* even if the service provider has no control over the occurrence of the event (except a change of control).[293] The legislative history provides,

> Amounts payable upon the occurrence of an event are not treated as amounts payable at a specified time. For example, amounts payable when an individual attains age 65 are payable at a specified time, while amounts payable when an individual's child begins college are payable upon the occurrence of an event.[294]

Section 409A also imposes the Triple Tax if the NQDC document provides that "assets will become restricted to the provision of benefits under the plan in connection with a change in the employer's financial health."[295] This reverses a previous IRS letter ruling which held that NQDC benefit payments could be triggered upon the employer's net worth falling below $10 million.[296]

2. The "No Acceleration" Rule

As recommended by the Joint Tax Committee in the Enron Compensation Report,[297] Section 409A effectively prohibits an acceleration of NQDC

---

290. In each situation, the NQDC plan would provide for distributions upon an event that is not an authorized trigger under Section 409A(a)(2)(A). As a result, the Triple Tax would apply because the plan does not satisfy the exact conditions of Section 409A(a)(2). I.R.C. § 409(A)(a)(1)(A)(i)(I).

291. *Id.* § 409A(a)(2)(A)(iv).

292. *See supra* notes 204-09 and accompanying text.

293. This approach is contrary to the established principles of the constructive receipt doctrine. *See supra* notes 183-203 and accompanying text.

294. Conference Agreement, *supra* note 13, at 730.

295. I.R.C. § 409A(b)(2)(A).

296. I.R.S. Priv. Ltr. Rul. 95-08-014 (Nov. 22, 1994) (concluding that the existence of that payment trigger would not cause the plan participants to be taxed on the NQDC benefits before the amounts were actually paid to the participants). The actual plan document considered in that ruling was reprinted. *See supra* note 204.

297. *See supra* note 143 and accompanying text.

payments through its imposition of the Triple Tax.[298] Minor exceptions to the "no acceleration rule" likely will be provided for (i) a distribution of a participant's entire account balance, if the account balance is under $10,000 and the participant separates from service, and (ii) other limited situations.[299]

---

298. I.R.C. § 409A(a)(3) ("The requirements of this paragraph are met if the plan does not permit the acceleration of the time or schedule of any payment under the plan, except as provided in regulations by the Secretary of the Treasury."). Section 409A(a)(1)(A)(i) provides that if a NQDC plan fails to meet the requirements of Section 409A(a)(2), (3) and (4), the Triple Tax applies.

299. The legislative history indicates that regulations to be issued by the Secretary of the Treasury should exclude "automatic distributions of minimal interests in a [NQDC] plan upon *permissible distribution events* for purposes of administrative convenience. For example, a plan could provide that upon separation from service of a participant, account balances less than $10,000 will be automatically distributed . . . ." Conference Agreement, *supra* note 13, at 732 (emphasis added). The exception for account balances under $10,000 will *only* be available if "the payment is made on or before the later of (A) December 31 of the calendar year in which occurs the participant's *separation from service* from the service recipient or (B) the date 2 1/2 months after the participant's *separation from service* from the service recipient . . . ." I.R.S. Notice 2005-1, 2005-2 I.R.B. 283 (emphasis added); *see* Prop. Treas. Reg. § 1.409A-3(h)(2)(iv)(A)(2), 70 Fed. Reg. 57930, 57980 (Oct. 4, 2005). Thus, the I.R.S. Notice could imply that the option to accelerate amounts less than $10,000 will be available only in connection with a separation from service and not in connection with any other permitted triggering events, such as a change of control.

Other exceptions under the proposed Treasury Regulations include: (i) distributions to comply with court-ordered settlements incident to a divorce; (ii) distributions required under Federal conflict of interest requirements; (iii) distributions to provide withholding of an employee's share of employment taxes; or (iv) distributions to an employee for payment of income taxes due upon a vesting event subject to section 457(f). Prop. Treas. Reg. § 1.409A-3(h)(2), 70 Fed. Reg. 57930, 57980 (Oct. 4, 2005); Conference Agreement, *supra* note 13, at 731; *see also* I.R.S. Notice 2005-1, 2005-2 I.R.B. 282-83 (discussing all the exceptions to the no acceleration rule). The last exception would be needed when amounts become taxable to a participant in a tax-exempt organization's NQDC plan described in Section 457(f) before the NQDC amount is actually paid to the participant. For example, if a nonprofit employer establishes a Section 457(f) NQDC Plan and an employee's benefits are subject to a substantial risk of forfeiture that subsequently expires, the employee will be subject to income tax on the NQDC benefits immediately, even though the benefits are not yet payable. *See* GOLDSTEIN, *supra* note 84, at 190-94 (discussing Section 457(f) plans established by tax-exempt entities). The proposed regulation also provides an exception for payments to cover FICA taxes, and income taxes incurred as a result of the distribution to pay FICA taxes. Prop. Treas. Reg. § 1.409A-3(h)(2)(v), 70 Fed. Reg. 57930, 57980 (Oct. 4, 2005); I.R.S. Notice 2005-1, 2005-2 I.R.B. 283. This exception is needed because vested NQDC is subject to FICA tax immediately, *see* I.R.C. § 3121(v)(2), even though the income tax is deferred. *See* GOLDSTEIN, *supra* note 84, at 79-80.

3. Prohibiting Subsequent Deferrals Shorter Than Five Years

Section 409A also removes flexibility of NQDC payments by prohibiting "subsequent elections" unless the first payment date is delayed at least five years.[300] While the Enron Compensation Report criticized flexibility in the company's NQDC plans, it fails to indicate that Enron and its top executives actually *delayed* NQDC payments after an original deferral.[301] Instead, the IRS' frustrations over a string of court decisions soundly rejecting its attempts to prohibit subsequent deferrals likely inspired this Section 409A restriction.

In the landmark *Veit II* case, Veit originally deferred a portion of his compensation from 1940 so that it would be paid to him in quarterly installments in 1942.[302] On December 26, 1941, Veit and his employer entered into a new agreement providing that the sum would be paid to Veit in five annual installments, beginning in 1943.[303] Despite the IRS' argument that Veit had constructively received the money because the subsequent deferral was made only five days before the end of the tax year, the Tax Court held that the agreement was effective to defer Veit's income tax liability on the benefit.[304] The IRS later lost other subsequent deferral cases in 1953 and 1991.[305]

Section 409A effectively reverses the holdings in these three cases[306] by providing that a subsequent election to delay a payment or change the form of a payment will trigger the Triple Tax unless:

> (i) *the plan requires* that such election may not take effect until at least 12 months after the date on which the election is made,
> (ii) . . . *the plan requires* that the first payment [impacted by the subsequent election must] be deferred for a period of not less than 5 years . . . , and
> (iii) [in the case of a payment scheduled to be made at a specified time or pursuant to a fixed schedule] *the plan requires* that any election . . . may

---

300. I.R.C. § 409A(a)(4)(C). The legislative history indicates Congress's intent that a subsequent election which fails to satisfy the requirements of Section 409A(a)(4)(C) will trigger the Triple Tax. Conference Agreement, *supra* note 13, at 733 ("A [NQDC] plan may allow a subsequent election to delay the timing or form of distribution *only* if [the plan meets the three requirements of Section 409A(a)(4)(C)].").

301. Instead, Enron and its executives were criticized for accelerating distributions.

302. *Veit II*, 8 T.C.M. (CCH) 919, 921-22 (1949). Veit deferred $87,076 of his 1940 salary, which was originally to be distributed in four installments of $21,769. *Id.*

303. *Id.* at 922.

304. *Id.* ("Under existing contracts there was never a time when the $87,076.40 was unqualifiedly subject to [Veit]'s demand or withdrawal.").

305. *See* Oates v. Comm'r, 18 T.C. 570, 585 (1952), *aff'd* 207 F.2d 711 (7th Cir. 1953), *acq.* 1960-2 C.B. 3; Martin v. Comm'r, 96 T.C. 814, 830 (1991).

306. These cases still stand for the proposition that the employee is not subject to income tax under the constructive receipt doctrine on amounts that are subsequently deferred. But that is little comfort if the subsequent election will cause the employee to be subject to the Triple Tax of Section 409A.

> not be made less than 12 months prior to the date of the first scheduled payment . . . .[307]

The subsequent election rule may come to be known as the "One Year/Five Year Rule" because the election must be made at least one year before the first scheduled payment, and the plan must delay the first payment for at least five years. The statute indicates that a subsequent election under the One Year/Five Year Rule may only be made if expressly provided for in the NQDC plan document.[308]

4. Other Section 409A Changes

Section 409A includes several other changes that will further complicate the drafting and administration of NQDC plans.

a. Rules on Elections to Defer Compensation

Presumably in order to make NQDC even more inflexible, under Section 409A(a)(4)(B), a NQDC plan must provide that a voluntary election to defer may only be made before the beginning of the year. But there are two exceptions: first, an election to defer can be made during a calendar year if the individual makes the election within thirty days of becoming eligible to participate in the plan;[309] second, if the deferred compensation qualifies as "performance-based compensation based on services performed over a period of at least 12 months, such election may be made no later than 6 months before the end of the period."[310]

b. Increased Reporting of NQDC on Forms W-2 and 1099

Even prior to the enactment of Section 409A, employers were required to include all amounts deferred (and which were vested) on employees' IRS W-2 Forms because those amounts were subject to FICA taxes.[311] But "earnings"

---

307. I.R.C. § 409A(a)(4)(C) (2005) (emphasis added).

308. Each of the three requirements under Section 409A(a)(4)(C)(i)-(iii) includes the phrase "the plan requires."

309. I.R.C. § 409A(a)(4)(B)(ii).

310. *Id.* § 409A(a)(4)(B)(iii). The legislative history provides that "'performance-based compensation' will be defined . . . to include compensation to the extent that an amount is: (1) variable and contingent on the satisfaction of preestablished organizational or individual performance criteria and (2) not readily ascertainable at the time of the election." Conference Agreement, *supra* note 13, at 732. As a result, if the employer agrees to pay a bonus contingent on performance from January through December, the employee can elect to defer all or part of the bonus as late as June 30th (if permitted by the NQDC plan document).

311. Section 3121(v)(2) of the I.R.C. includes within the definition of wages for FICA tax purposes, nonqualified deferred compensation at the later of (i) when the services are performed, or (ii) when there is no substantial risk of forfeiture of the rights to such amount.

on NQDC amounts are not subject to FICA tax,[312] and prior to the effective date of Section 409A, were not required to be reported on an employee's Form W-2. Amounts deferred but subject to a "substantial risk of forfeiture" are not subject to FICA tax[313] and, prior to the effective date of Section 409A, were not required to be reported on an employee's Form W-2.

For amounts deferred after December 31, 2004,[314] Section 409A(d)(1) amends the requirements for Form W-2 to require annual reporting of "the total amount of deferrals for the year under a nonqualified deferred compensation plan (within the meaning of section 409A(d))."[315] The practical impact of this change is that employers are now required to report annually: (i) all amounts deferred for the benefit of an employee, whether or not subject to a substantial risk of forfeiture; and (ii) earnings on amounts deferred after 2004. The second requirement may be particularly burdensome for many employers. Earnings on amounts deferred before 2005 are not subject to Section 409A and employers are not bound by this reporting requirement,[316] but earnings on amounts deferred after 2004 are subject to Section 409A and this reporting requirement.[317] Consequently, to prepare the annual Form W-2 for each employee, the employer must allocate the earnings in the employee's NQDC account balance between the pre-2005 amounts, and the post-2004 amounts.[318] Similarly, when NQDC benefits are provided for an independent contractor, Section 409A(b)(3) now requires that all NQDC amounts subject to Section

---

I.R.C. § 3121 (v)(2)(A). There are two types of FICA taxes: (i) the Old Age, Survivor and Disability Insurance ("OASDI") portion (sometimes referred to as the " Social Security Tax"); and (ii) the Medicare Tax. GOLDSTEIN, *supra* note 84, at 78. The OASDI tax rate is 6.2% on the employer, and 6.2% on the employee, but only applies up to an applicable wage ceiling for each employee, each year. *Id.* For 2005, the wage ceiling is $90,000 of wages. 42 U.S.C. § 430 (2000). Thus, if the employee receives wages in excess of $90,000 for the year, the NQDC amounts will not increase the employee's (or the employer's) OASDI tax burden. In contrast, since 1993 there has been no "cap" on the Medicare Tax. *See* GOLDSTEIN, *supra* note 84, at 78. The Medicare Tax applies to all NQDC amounts deferred during a year that are not subject to a substantial risk of forfeiture. The Medicare Tax rate is 1.45% on the employee, and 1.45% on the employer. *See generally* GOLDSTEIN, *supra* note 84, at 77-101 (discussing the application of FICA taxes to NQDC).

312. I.R.C. § 3121(v)(2)(B) ("Any amount taken into account as wages . . . (and the income attributable thereto) shall not thereafter be treated as wages for [FICA tax purposes]."); *see* GOLDSTEIN, *supra* note 84, at 89.

313. I.R.C. § 3121(v)(2). These amounts are not subject to FICA tax until the risk lapses.

314. American Jobs Creation Act of 2004, Pub. L. No. 108-357, § 885(d)(1), 118 Stat. 1418 (2004).

315. *Id.* § 885(b)(1)(A).

316. This assumes that there is no material modification of the NQDC plan after October 3, 2004. *See supra* notes 249-58 and accompanying text.

317. American Jobs Creation Act of 2004 § 885(b)(3).

318. The IRS may need to issue guidance advising employers on the method for making this allocation in the case of "defined benefit" NQDC plans.

409A be reported on the IRS Form 1099 provided to the independent contractor.[319]

D. Assessing the Impact of Section 409A

The enactment of Section 409A will impose significant transaction costs on employers and employees in the short term. In almost all cases, to preserve the grandfathered status of pre-2005 deferrals and protect the employees from the Triple Tax, existing NQDC plans must have been amended by the end of 2005 (or another deadline to be specified by the IRS).[320] Almost every employer maintaining a NQDC arrangement will need to arrange for an expert to review and revise the plan to include all the magic language and other requirements of Section 409A. Not only has Section 409A imposed significant administrative burdens on employers in connection with Forms W-2 and 1099, it has also added substantial burdens on the Treasury Department and the IRS to issue guidance and regulations to interpret Section 409A. Further, beyond these short-term problems, perhaps one of the most disturbing aspects of Section 409A is that the sophisticated, well-informed employers and their employees will face minor irritation, while the employees of the unwary may fall into a dangerous trap that serves no constructive purpose.

1. Section 409A Will Be a Minor Irritation for the Well-Informed

Section 409A primarily establishes timing rules. In summary, Section 409A will prevent an employer from (i) establishing a NQDC plan that pays benefits other than upon the occurrence of one or more of the six Payment Triggers,[321] (ii) accelerating the payment of any NQDC,[322] (iii) extending the time for payment of any NQDC after the amount has already been deferred,[323]

---

319. *See* American Jobs Creation Act of 2004 § 885(b)(3) (amending I.R.C. § 6041(g)); I.R.S. Notice 2005-1, 2005-2, I.R.B. 288-89.

320. While it is theoretically possible that a plan drafted before the enactment of Section 409A would comply with 409A, this is extremely unlikely because of its magic language requirements. For example, a NQDC plan drafted before Section 409A could have provided that a participant would not be considered "disabled" until she has received at least three months of disability payments (as required by Section 409A(a)(2)(c)), but this seems extremely unlikely.

321. *See* I.R.C. § 409A(a)(2)(A)(i)-(vi) (2005). One of the permitted Payment Triggers is at a "specified time" (or pursuant to a fixed schedule) designated under the plan at the date of the deferral of the compensation. *Id.* § 409(a)(2)(A)(iv). Thus, an employer and employee have some flexibility at the time the compensation is initially deferred, but thereafter no variations are permitted.

322. *Id.* § 409A(a)(3).

323. *Id.* § 409A(a)(4)(C). An exception to this rule occurs when the subsequent amendment is made at least one year before the first payment is due, and extends the first scheduled payment for at least another five years. *Id.*

and (iv) allowing a participant to defer additional amounts once the taxable year has begun.[324]

Section 409A's timing rules accomplish little in the case of well-advised employers and employees other than forcing them to incur transaction costs to comply with these rules. According to the Enron Compensation Report, the "no acceleration rule" will prevent top executives from raiding a corporation's funds immediately before bankruptcy, to the detriment of the corporation's other employees, creditors, and shareholders.[325] While this purpose sounds noble, and commentators seem thrilled to make this argument (even when they acknowledge that it is fallacious),[326] the bankruptcy laws already prohibit top executives from profiting from such behavior.[327] But the rationale for the other timing rules is a mystery.[328]

Most importantly, Section 409A will impact neither the total amount of compensation paid to top executives of public companies, nor the motivations of directors in negotiating compensation arrangements with these top executives. The Triple Tax only applies when the timing of payments (either under the NQDC plan documents, or in the actual administration of the NQDC plan) fails to conform to the requirements of Section 409A. Top executives and their employers are free to defer the same amounts of compensation as before the enactment of Section 409A; they simply cannot amend the payment schedule once it has been established at the time of deferral, except for subsequent elections that satisfy the One Year/Five Year Rule.[329] If any top executive views NQDC as less attractive because of the inflexibility required by Section 409A, they may simply choose to defer less of their compensation and receive more currently. Also, Section 409A will not encourage the board of directors of public companies to restrict NQDC benefits to top executives because it does not increase disclosure requirements to shareholders or the

---

324. *Id.* § 409A(a)(4)(B). Exceptions are provided if (i) an individual becomes eligible to participate in the plan after the beginning of the year, or (ii) the compensation is "performance-based." *Id.* § 409A(a)(4)(B)(ii)-(iii).

325. Enron Compensation Report, *supra* note 2, at 636 ("The Joint Committee Staff recommends that under [NQDC] arrangements, plan provisions allowing accelerated distributions at the request of the participant, should trigger constructive receipt rather than resulting in deferral. Distributions made to [Enron] executives . . . immediately preceding the bankruptcy drained Enron's cash by over $53 million that would have been available to the creditors . . . .").

326. *See* Diana B. Henriques & David Cay Johnston, *Managers Staying Dry as Corporations Sink,* N.Y. TIMES, Oct. 14, at A1 (acknowledging that top executives can only preserve a portion of their NQDC benefits through negotiations); Kathryn J. Kennedy, *A Primer on the Taxation of Executive Deferred Compensation Plans,* 35 J. MARSHALL L. REV. 487, 520 (2002) ("[T]he bankruptcy courts should have no problems treating these withdrawals as voidable and using the proceeds for Enron creditors.").

327. *See supra* Part II.F.1.

328. The legislative history of Section 409A fails to provide any rationale for these rules.

329. *See* I.R.C. § 409A(a)(4)(C).

general public.[330] Greater disclosure would increase the risk of public and shareholder outrage, but Section 409A merely requires additional disclosure about NQDC on IRS Form W-2 and IRS Form 1099, which are only available to the IRS. Thus, while these rules will require some changes to the NQDC documents and additional diligence in administering NQDC plans, the bottom line is that Section 409A likely will have no significant impact on the way the well-informed employers and excessively-compensated employees utilize NQDC arrangements.

2. The Impact of Section 409A on the Uninformed—Traps for the Unwary

While Section 409A may prove bothersome for the well-informed, it may devastate the service providers of uninformed employers. As a preliminary matter, the Triple Tax only falls on the service provider.[331] In contrast, violating Section 409A may actually *benefit* the employer by accelerating its income tax deduction for compensation provided to the employee.[332] A plan may violate Section 409A in numerous ways, but from an employee's perspective, an acceleration of the employee's entire account balance may be best because the employee will at least have the cash to pay the tax.[333] But the Triple Tax can be triggered in numerous situations when the employee receives none, or a very small portion, of her NQDC benefits.[334] The employee's liability for the Triple Tax could be substantial in the current year, even though she will not receive, and will not have access to, any of her NQDC benefits for many years. This section will summarize several potential land mines of Section 409A.

---

330. As discussed above, directors may argue more forcefully about executive compensation if excessive compensation may trigger public or shareholder outrage. *See supra* notes 24-36 and accompanying text.

331. I.R.C. § 409A(a)(1)(A)(II) (stating that the Triple Tax applies if the NQDC plan fails to satisfy Section 409A(a)(2), (3), or (4), or "is not operated in accordance with such requirements").

332. An employer may deduct NQDC when it is included in the employee's gross income, provided that the compensation is "reasonable" under Section 162(a)(1) of the I.R.C. I.R.C. § 404(a)(5). Thus, by violating Section 409A, an employer may accelerate its income tax deduction because the violation will accelerate when the employee is required to include the amount in her gross income.

333. For instance, in Example #4, *supra* text accompanying note 229, Tom Tragic is obligated to pay the Triple Tax of $89,821 for the 2005 tax year, but as a result of receiving his entire NQDC account balance in 2005, he received $142,068 in cash from the NQDC plan in 2006. Thus, Tom would have the cash to pay the tax.

334. For example, the plan document may fail to fulfill one of the magic language provisions of Section 409A, such as the definition of disability, separation from service or change of control. *See supra* Part IV.C.1.

a. Failing to Amend a NQDC Plan Established Before the IRS Deadline[335]

A NQDC plan established before 2005 will be deemed to have been materially modified if any additional amounts are deferred after 2004, or if any other additional "benefit, right or feature" is added after 2004.[336] As a result, the participants under these plans will be subject to the Triple Tax if the plan document is not properly amended to comply with Section 409A before the IRS deadline.[337] Additionally, if the plan is not *operated* in compliance with Section 409A for any reason, the employee will be subject to the Triple Tax.[338]

b. Creating New NQDC Plans After 2005

Amounts deferred after December 31, 2004 will be subject to Section 409A.[339] As such, an employee will be subject to the Triple Tax if either the plan document fails to satisfy the magic language or other requirements of Section 409A, or the employer fails to administer the arrangement according to the strict requirements of Section 409A. These rules are especially important because of the broad definition of deferred compensation.[340] A simple employment agreement that defers a bonus payment for more than two and a half months will create a deferral of compensation subject to Section 409A.[341] When a Section 409A violation is discovered quickly after the deferral, the primary tax costs to the employee will be the 20% "bonus" tax liability and the acceleration of taxation on the current year deferral.

---

335. *See supra* notes 245-54 and accompanying text.

336. Conference Agreement, *supra* note 13, at 737. But the mere accrual of earnings on amounts deferred before 2005 should not constitute a "material modification" for purposes of Section 409A.

337. Although it is theoretically possible that a plan document prepared before the enactment of Section 409A would comply with Section 409A, it is highly unlikely because of the "magic language" requirements of Section 409A. *See supra* Part IV.C.

338. For instance, in Example #10, the employer failed to operate the NQDC arrangement in compliance with Section 409A because Old Reliable received NQDC benefits upon switching from full-time to part-time employment. *See supra* notes 282-84 and accompanying text; *see also infra* note 362 and accompanying text (describing Example #13, which involves Nancy, an employee who requests a distribution to pay her child's tuition).

339. American Jobs Creation Act of 2004, Pub. L. No. 108-357, § 885(d)(1), 118 Stat. 1418 (2004).

340. *See* I.R.C. § 409A(d)(1) (2005) ("The term 'nonqualified deferred compensation plan' means any plan that provides for the deferral of compensation . . . ."); I.R.S. Notice 2005-1, 2005-2 I.R.B. 277 ("A plan provides for the deferral of compensation only if, under the terms of the plan and the relevant facts and circumstances, the service provider has a legally binding right during a taxable year to compensation that has not been actually and constructively received and included in gross income . . . .").

341. I.R.S. Notice 2005-1, 2005-2 I.R.B. 277-78.

EXAMPLE #11. Uninformed Employer hires the law firm of Quick & Cheap to prepare an employment agreement for a key employee, Unlucky Lucia. The lawyers at Quick & Cheap are unaware of Section 409A and provide in her employment agreement that: (i) in addition to her regular salary Lucia will receive a $30,000 "bonus" on April 1 of Year 2 for her services during Year 1; (ii) Lucia is entitled to the bonus if she is employed on June 30 of Year 1; and (iii) the bonus will be distributed immediately if Lucia becomes disabled. The employment agreement triggers the Triple Tax because it fails to include the three-month rule of Section 409A(a)(2)(C) in the definition of "disability."[342] As a result, Lucia likely will incur an additional income tax liability of approximately $16,000 in Year 2.[343]

The Triple Tax may create an even greater economic hardship for the employee if compensation has been deferred for many years and the violation is not discovered for many years.

The statute of limitations will provide no relief to the employee in these situations. As a general rule, income tax may only be assessed within three years of the due date of the tax return.[344] But the Triple Tax creates a tax liability in the current year, even if the violation of Section 409A occurred in taxable years that otherwise would be closed under the applicable statute of limitations.[345]

c. Economic and Tax Problems if the Benefit Is Never Paid

Section 409A may create situations in which employees are forced to pay income tax on amounts they will never receive, which is not an unprecedented income tax result. Since the "constructive receipt doctrine" triggers a tax on amounts that have been constructively received, but not actually received,[346] the individual will be taxed on an amount that she never receives if the obligor is unable to pay. Similarly, under the "claim of right doctrine," taxpayers must pay tax when they actually receive property, even if they subsequently must

342. The NQDC arrangement includes a Payment Trigger other than the six magic Payment Triggers permitted by Section 409A(a)(2)(A). The bonus is not subject to a substantial risk of forfeiture after June 30. *See* I.R.C. § 409A(1)(a)(i)(II).

343. In this situation, the major tax costs to Lucia will be the 20% additional tax liability of Section 409A(a)(1)(B)(i)(II), which results in additional tax of $6,000 [$30,000 x 20% = $6,000], and the tax on the acceleration of the Year 2 bonus. If Lucia is in the 35% federal income tax bracket, the tax on $30,000 would be $10,500 [$30,000 x 35% = $10,500]. If the bonus was not taxed under Section 409A in Year 1, Lucia would have to pay tax on the $30,000 when she received it in Year 2.

344. I.R.C. § 6501(a).

345. *Id.* § 409A(a)(1)(A)(i)(II) ("[A]ll compensation deferred under the plan for the taxable year *and all preceding taxable years* shall be includible in gross income for the taxable year to the extent not subject to a substantial risk of forfeiture and not previously included in gross income." (emphasis added)).

346. Treas. Reg. § 1.451-2(a) (2005).

return the property received.[347] For example, an embezzler must pay tax in the year he embezzles the funds, even if he is obligated to make restitution.[348]

In a claim of right situation, the taxpayer may be entitled to a tax deduction upon repayment of the property.[349] If the individual includes an amount in taxable income in Year 1 as a result of receiving the property, and must repay in Year 2, he may be entitled to a tax deduction in Year 2. If the taxpayer is taxed at the same rate in Years 1 and 2, presumably the only economic cost to the taxpayer will be the time value of the money. If the individual in a claim of right situation is not a scoundrel, section 1341 of the I.R.C. may even be available to protect the taxpayer from differences in tax rates between the year the taxpayer received the property and the year he paid it back.[350] If employees pay tax on NQDC under the constructive receipt doctrine, and never receive the amount, presumably they will be entitled to a bad debt deduction when their right to receive the amount becomes worthless.[351]

---

347. *See* James Edward Maule, *Gross Income: Tax Benefit, Claim of Right and Assignment of Income*, *in* 502-2d TAX MANAGEMENT PORTFOLIOS, at A-23 (BNA 2003) ("Under the claim of right doctrine, a taxpayer who receives any income under a claim of right that is free of restrictions must include that income in gross income for the year of receipt. It does not matter that the taxpayer is not entitled to retain the income and might be obligated to return it." (footnote omitted)).

348. *See* James v. United States, 366 U.S. 213, 221 (1961); Maule, *supra* note 347, at A-28.

349. North Am. Oil Consol. v. Burnet, 286 U.S. 417, 424 (1932); Maule, *supra* note 347, at A-33.

350. I.R.C. § 1341(a)(1) ("If . . . it appeared that the taxpayer had an unrestricted right to such item; a deduction is allowable."). *See* McKinney v. United States, No. A-74-CA-239, 1976 WL 1151, at *3 (W.D. Tex. Sept. 30, 1976) ("It seems almost too clear for argument that funds received through embezzlement . . . are not held under a claim of right or with the appearance that the embezzler has an unrestricted right to the use of the embezzled funds.").

351. I.R.C. § 166(a)(1) ("There shall be allowed as a deduction any debt which becomes worthless within the taxable year."). In *Patrick v. Comm'r*, 14 T.C.M. (CCH) 234, 234-35 (1955) the IRS argued that the taxpayer constructively received amounts held by his wholly-owned corporation (in his "personal account") under the constructive receipt doctrine. *Id.* The taxpayer argued that if he was subject to income tax on the amount under the constructive receipt doctrine, he was entitled to an income tax deduction under the predecessor statute of section 166 of the I.R.C. because the corporation experienced financial difficulties and never paid the amounts to him. *Id.* The court concluded that the taxpayer was never in constructive receipt of the amounts held by the corporation, and as a result, the court never reached the issue of deductibility. *Id.* at 235. A potential issue would be whether the bad debt was incurred in a trade or business (in which case the bad debt deduction could offset the employee's ordinary income) or was a personal bad debt which can only be used to offset capital gains, and up to $3,000 of ordinary income per year. *See* I.R.C. § 166(d)(1)(B) ("In the case of a taxpayer other than a corporation . . . where any nonbusiness debt becomes worthless within the taxable year, the loss resulting therefrom shall be considered a loss from the sale or exchange . . . of a capital asset held for not more than 1 year."). 2 BORIS BITTKER & LAWRENCE LOKKEN, FEDERAL TAXATION OF INCOME, ESTATES AND GIFTS ¶ 46.2.5 (3d ed. 2000) ("Under [I.R.C.] § 1211(b), an individual . . . may deduct all capital losses (whether short or long term) against all capital gains

Section 409A, however, creates a bigger tax problem for the unlucky employee. Under the Triple Tax, the employee not only pays tax on the deferred amounts, but also pays an additional tax "equal to 20 percent of the compensation which is required to be included in gross income" under Section 409A.[352] Under the constructive receipt doctrine, if the employee never receives the deferred amounts, the employee may be entitled to a deduction equal to the amounts deferred. But there appears to be no mechanism entitling the employee to receive a refund of the additional 20% tax.

> *EXAMPLE #12.* The facts are the same as in Example #11, but Uninformed Employer declares bankruptcy in Year 2, and Unlucky Lucia never receives the $30,000 bonus. She paid $6,000 [$30,000 x 20%] in extra income tax because of the 20% additional tax of Section 409A(a)(1)(B)(i)(II), but there appears to be no mechanism for Lucia to recover that extra tax (even though she never received the $30,000).

It is particularly ironic that this problem is not addressed in Section 409A because this problem would have presented itself for the Enron executives if Section 409A had been enacted.[353]

. . . . If the capital losses in the aggregate exceed the capital gains, up to $3,000 of the excess can be deducted against ordinary income . . . ."). Employees generally may deduct employee business expenses that have not been reimbursed only as itemized deductions. *See* I.R.C. § 62(a)(1) (allowing a deduction against gross income in computing adjusted gross income for trade or business expenses unless the trade or business consists of "the performance of services by the taxpayer as an employee"); James E. Maule & Lisa M. Starczewski, *Deductions: Overview and Conceptual Aspects*, *in* 503-2d TAX MANAGEMENT PORTFOLIOS, at A-18 (BNA 2000) ("Employee trade or business expenses that are not reimbursed are allowable as deductions from adjusted gross income . . . subject to limitations such as the § 67 limitation on miscellaneous itemized deductions . . . .").

352. I.R.C. § 409A(a)(1)(B)(i)(II).

353. In response to Enron's executive compensation abuses, the Enron Compensation Report suggests most of the rules adopted in Section 409A. *See supra* notes 141-45 and accompanying text. Specifically, if Section 409A had applied, Enron's December 2001 distributions would have constituted accelerations in violation of Section 409A(a)(3), triggering the Triple Tax, including the additional 20% tax. In reality, the Enron bankruptcy trustee can recover the NQDC amounts that were distributed to the top executives immediately before bankruptcy. Thus, those Enron executives would have paid an extra 20% tax on money they would have to repay.

V. FUNDAMENTAL PROBLEMS WITH SECTION 409A THAT MERIT RETROACTIVE CORRECTION

The rules of Section 409A are difficult to justify,[354] as they appear to serve no necessary purpose.[355] If Section 409A discourages the deferral of compensation, it will encourage the payment of more current compensation.[356] Thus, it may discourage savings and increase consumption. This runs counter to efforts to increase savings and reduce compensation for economic reasons, such as to encourage economic growth. Certain features of Section 409A are especially bizarre and merit prompt revision or repeal. Such drastic measures are not unprecedented; in previous years, Congress has retroactively repealed tax rules that posed significant problems.[357]

A. The Imposition of the 20% Additional Tax in Many Situations Will Be Ridiculous

While fine-tuning remedies for different behaviors necessarily adds complexity, it can reduce the potential for unjust results. Section 409A demonstrates the absurd results that can occur when the same remedy is applied to any behavior that falls short of the ideal. Section 409A punishes with a sledge-hammer, even when the problem is as small as an ant.

---

354. At least two of the minor rules are supportable. First, the requirement to report earnings on NQDC amounts on Forms W-2 and 1099, *see supra* Part IV.C.4.b., will provide the IRS with more complete information about the use of NQDC, which may lead to improved rule-making in the future. Second, in order to eliminate an improper advantage available to those who use off-shore trusts to shield assets from the claims of the employer's creditors, a provision of Section 409A is designed to immediately tax those benefits. *See* I.R.C. § 409A(b)(1). Other Section 409A rules can be viewed as simply providing clarification—for example, the addition of a clear definition of "change of control," *see supra* Part IV.C.1.c., will provide guidance to practitioners when drafting a change of control payment trigger. But instead of creating mandatory magic language to avoid the Triple Tax, creating "safe harbors" may have been a more appropriate tool to aid service providers and recipients.

355. The Enron Compensation Report asserts that the purpose for the no acceleration rule is to prevent the distribution of NQDC to top executives immediately before the company goes bankrupt, to the disadvantage of other employees, creditors, and shareholders. *See supra* notes 128-44 and accompanying text. But such a problem does not even exist because the bankruptcy trustee can recover those funds for the benefit of the bankruptcy estate. *See supra* Part II.F.1.

356. Presumably, if the rules for deferring compensation are more onerous, some employers and employees will discontinue deferral plans, and amounts that otherwise would be deferred will be distributed immediately to the participants.

357. For example, the Tax Reform Act of 1986, Pub. L. No. 99-514, § 1433(c), 100 Stat 2731 (1986), retroactively repealed the generation-skipping tax (chapter 13 of the I.R.C.), and created a new generation-skipping tax effective for transfers made on or after October 22, 1986. *See* Carol A. Harrington & Frederick G. Acker, *Generation-Skipping Tax*, *in* 850 TAX MANAGEMENT PORTFOLIOS, at A-70 (BNA 1996).

The enforcement mechanism of Section 409A is the Triple Tax. The 20% additional tax feature of the Triple Tax was probably designed with accelerations in mind. When an acceleration, also referred to as a "premature distribution," is made from a qualified plan or an IRA before the participant attains age 59-1/2, an additional tax equal to 10% of the amount withdrawn is imposed (with certain exceptions).[358]

1. The 20% Additional Tax Will Apply to Certain Premature Distributions that Should Not Be Penalized

Section 409A establishes only six Payment Triggers.[359] Distribution at any other time will subject the employee to the Triple Tax. The exceptions from the 10% early withdrawal penalty under the qualified plan rules are more liberal and appear to have been more thoroughly considered. For example, Section 409A's Payment Trigger for "unforeseeable emergency" is available only for "extraordinary and unforeseeable circumstances arising as a result of events beyond the control of the participant."[360] A distribution to a parent to help pay a child's tuition would generally not be permitted.[361]

> *EXAMPLE #13.* Qualified Quentin's employer established a 401(k) plan for his benefit. When he terminated employment, Quentin rolled-over his $100,000 account balance to an IRA. Nonqualified Nancy's employer established a NQDC plan for her benefit, and her NQDC balance is $100,000. Even though Quentin has not yet attained age 59-1/2, he can withdraw from his IRA to help pay his son's tuition without triggering the 10% early withdrawal penalty.[362] But if Nancy's employer distributes NQDC amounts to Nancy to help her pay for her daughter's tuition, Nancy will be subject to the Triple Tax of Section 409A, assuming no Payment Trigger described in Section 409A(a)(2)(A) (i)-(vi) otherwise has occurred.

2. The 20% Additional Tax Applies to All Amounts Deferred, Not Just the Amount Distributed

Congress apparently concluded that a premature distribution from a NQDC plan is more evil than an early distribution from a qualified plan because it

---

358. I.R.C. § 72(t)(1)-(2) (providing that the 10% additional tax does not apply if the distribution is "made on or after the date on which the employee attains age 59-1/2").

359. *See supra* Part IV.C.1.

360. I.R.C. § 409A(a)(2)(B)(ii)(I).

361. *See* Treas. Reg. § 1.457-6(c)(2)(i) (2005) ("Except as otherwise specifically provided in this paragraph . . . the purchase of a home and the payment of college tuition are not unforeseeable emergencies . . . .").

362. *See* I.R.C. § 72(t)(2)(E), (7)(A)(iii) (allowing distributions to pay for the "qualified higher education expenses" of the taxpayer's child or grandchild); Kathryn J. Kennedy, *IRAs, SEPs and SIMPLEs*, *in* 355-6th TAX MANAGEMENT PORTFOLIOS, at A-35 (BNA 2004).

imposed a higher additional tax on a premature distribution from a NQDC plan (20% instead of 10%).[363] The Section 409A 20% tax applies to *all amounts deferred* for the benefit of the employee for all years even if only a de minimis amount is withdrawn.[364] In contrast, the 10% penalty applicable to an early distribution from a qualified plan only applies to the amount actually distributed.[365] Furthermore, the 10% haircut provision, as used in the Enron plan, permits the *employer* to keep 10% of the amount requested and distributed, but the 20% additional tax funnels 20% of the *total amount deferred* to the IRS, without changing the assets or liabilities of the employer who kept 10%.[366]

3. The 20% Additional Tax Applies to Any Section 409A Violation, Not Just Premature Distributions

While the 20% additional tax may have been designed to increase the economic cost of a premature distribution,[367] it applies to any Section 409A violation.[368] The application of the 20% additional tax to other Section 409A violations is bizarre. For example, assume that a NQDC agreement provides for distributions upon disability, and the employee will be considered disabled after receiving payments from the employer's long-term disability plan for two months. Since the plan uses "two months" instead of "three months" in its definition of disabled, the plan violates Section 409A.[369] As a result, the employee will be subject to an extra tax equal to 20% of all amounts deferred. One may find it difficult to explain to the employee why she is immediately subject to income tax on all the amounts deferred, plus a 20% additional tax on all amounts deferred, even though she may not receive the NQDC benefits for twenty or thirty years. Arguments based on ability to pay or constructive receipt would not be particularly persuasive. The employee's right to receive a benefit in twenty or thirty years, or to become disabled and begin to receive NQDC payments after two months, hardly seems "as good as cash," and will not give the employee the ability to pay. Even if the government believes that the disability distribution benefit is as good as cash, which appears to be the

---

363. As Section 409A worked its way through Congress, the additional tax was increased from zero in the House, Conference Agreement, *supra* note 13, at 722, to 10% in the Senate, *id.* at 727, to 20% in the Joint House-Senate Conference Agreement. *Id.* at 734.

364. I.R.C. § 409A(a)(1)(B)(i)(II). This aspect of Section 409A was illustrated in Example #4 of this Article, in which Tom Tragic receives an early distribution of $5,000, and owes $89,821 in taxes as a result.

365. *Id.* § 72(t)(1).

366. *See supra* notes 128-30 and accompanying text.

367. *See supra* Part V.A.1-2.

368. I.R.C. § 409A(a)(1).

369. *Id.* § 409A(a)(2)(C)(ii).

approach taken in the House Report, imposing tax and interest on the deferral amount, but no additional 20% tax,[370] would be preferable to the Triple Tax.

4. In the Case of Drafting Errors, the 20% Additional Tax Will Apply to All Plan Participants

Section 409A unjustifiably imposes the Triple Tax for "foot faults,"[371] minor errors that provide no significant economic benefit to the employee. And in the case of a drafting error, the Triple Tax will be imposed on all plan participants, as it applies to "all compensation deferred under the plan *for participants with respect to whom the failure relates*."[372] Admittedly, the number of employees who can participate in an employer's NQDC plan will be limited to a "select group of management or highly compensated employees."[373] Nevertheless, courts customarily have held that up to 5% of an employer's work force can be covered by a NQDC plan;[374] one case even suggests that up to 15% of the employer's work force can be covered by a NQDC plan.[375]

---

370. *See* Conference Agreement, *supra* note 13, at 722.

371. In the context of executives behaving badly, using the term "foot fault" to describe a challenged action may have the same consequences as waving a red flag in front of a charging bull. In a now infamous exchange, Hank Greenberg, the former chairman and chief executive officer of AIG, referring to the investigation by New York Attorney General Elliott Spitzer's office, said,

"When you begin to look at foot faults and make them into a murder charge, then you have gone too far."

> That evening Spitzer, who had read about Greenberg's remarks on the Internet . . . departed from his speech [at a gathering] to address the AIG chief's comments. "Hank Greenberg should be very, very careful talking about foot faults," he warned. "Too many foot faults, and you can lose the match. But more importantly, these aren't just foot faults."

Devin Leonard, *All I Want in Life Is an Unfair Advantage*, FORTUNE, Aug. 8, 2005,at 76; *see also A.I.G.'s View of Regulators*, N.Y. TIMES, Aug. 12, 2005, at C7 (discussing the recently ousted Greenberg and his famous "foot faults" quote).

372. I.R.C. § 409A(a)(1)(A)(ii) (emphasis added). The legislative history provides the following example:

> [S]uppose a plan covering all executives of an employer . . . allows distributions to individuals subject to section 16(a) [of the Securities and Exchange Act of 1934] upon a distribution event that is not permitted under [Section 409A]. The individuals subject to section 16(a), rather than all participants of the plan, would be required to include amounts deferred in income and would be subject to interest and the 20-percent additional tax.

Conference Agreement, *supra* note 13, at 729-30.

373. ERISA § 201(2), 29 U.S.C. § 1051(2) (2000); ERISA § 301(a)(3), 29 U.S.C. § 1081(a)(3); ERISA § 401(a)(1), 29 U.S.C. § 1101(a)(1); *see* GOLDSTEIN, *supra* note 84, at 132-38 (stating that eligible individuals frequently are referred to as the "top hat" group).

374. The 5% "rule of thumb" likely can be traced to *Belka v. Rowe Furniture Corp.*, 571 F. Supp. 1249, 1252 (D. Md. 1983), in which 4.6% of the employer's workforce was covered by a NQDC agreement. The court concluded that the NQDC was "established for a group of highly paid or management-level employees," and thereby qualified as a top hat plan that was not

EXAMPLE #14. Big Co. has 5,000 employees, 250 of whom participate in Big Co.'s NQDC plan. In 2005, Big Co.'s attorneys prepare an amendment to the NQDC plan providing that deferred amounts will be paid upon disability, and a participant is deemed to be disabled after receiving two months of long-term disability benefits. Because the permitted Payment Trigger for disability under Section 409A only allows distributions after the participant has been receiving disability benefits for three months,[376] this technical drafting error causes the NQDC plan to violate Section 409A. A few years later, an IRS agent spots the drafting error, yells, "Gotcha," and asserts the Triple Tax against all 250 plan participants.[377]

NQDC plans are aggregated for purposes of Section 409A.[378] Thus, if there is a Section 409A violation with respect to one plan, the NQDC plans maintained for all of those participants will be deemed to violate Section 409A, and the Triple Tax will apply to all amounts deferred under all the plans. Some might argue that participants should not be entitled to their NQDC benefits after receiving long-term disability payments for only two months; but even these individuals may agree that the immediate imposition of the 20% additional tax upon all participants on their deferred and unpaid NQDC amounts is a disproportionate response.

B. Distributions Upon the Occurrence of Events with Independent Significance Should Not Trigger the Triple Tax

The rationale for the rule that distributions can be made only upon the occurrence of the six Payment Triggers specified in Section 409A, and no other events, is a mystery. Perhaps it was viewed as simply a way to make NQDC more inflexible, as suggested by the Enron Compensation Report.[379] The legislative history states:

> Amounts payable at a specified time or pursuant to a fixed schedule must be specified under the plan at the time of deferral. Amounts payable upon the

---

subject to the more burdensome requirements of ERISA. *See id.* at 1253; GOLDSTEIN, *supra* note 84, at 135.

375. Demery v. Extebank Deferred Comp. Plan (B), 216 F.3d 283, 289 (2d Cir. 2000).

376. I.R.C. § 409A(a)(2)(C)(ii). Distributions would also be permitted if the participant is "unable to engage in any substantial gainful activity," but that would be a very difficult standard to satisfy. *Id.* § 409A(a)(2)(C)(i); *see also supra* note 276 (discussing this definition as it applies in other statutes unrelated to Section 409A).

377. *See* Mogab v. Comm'r, 70 T.C. 208, 214 (1978) (providing an example of the IRS using technical requirements as a "gotcha" rule). In *Mogab,* the IRS denied the taxpayer a deduction because the corporation's plan document for issuing stock described in section 1244 of the I.R.C. failed to specify the maximum dollar amount to be received under the plan. *Id.* at 211, 214.

378. I.R.C. § 409A(d)(6).

379. *See supra* notes 141-45 and accompanying text.

> occurrence of an *event* are not treated as amounts payable at a specified time. For example, amounts payable when an individual attains age 65 are payable at a specified time, while amounts payable when an individual's child begins college are payable upon the occurrence of an *event*.[380]

A fundamental tax principle is that amounts should become taxable when the taxpayer has the ability to pay.[381] Under the constructive receipt doctrine, amounts need not be included in taxable income when the taxpayer's ability to obtain the funds is subject to a substantial limitation or restriction,[382] such as the occurrence of an event that the taxpayer does not control.[383] The following payment triggers should not trigger immediate taxation: (i) a child or grandchild incurring tuition expenses;[384] (ii) a relative incurring medical expenses; (iii) a parent incurring nursing home costs; or (iv) involuntarily switching from full-time to part-time employment.[385]

While Section 409A allows distributions upon the "occurrence of an unforeseeable emergency," the definition of "unforeseeable emergency" is extremely narrow.[386] The rules for 401(k) plans and other qualified plans, however, are much more liberal on when participants can receive distribution without triggering additional taxes or penalties (other than the tax on the amount actually distributed).[387] It would seem reasonable to at least allow distributions at the times permitted for withdrawals from 401(k) plans and other qualified plans.

C. The Bizarre "Subsequent Deferral" Rule

The Triple Tax will apply if a NQDC plan "permits . . . a *subsequent election* [to] delay a payment or a change in the form of the payment"[388] where (i) the election takes effect at least one year after the date the election was made, and (ii) the first scheduled payment will be deferred for at least five years.[389]

---

380. Conference Agreement, *supra* note 13, at 723 (emphasis added).

381. *See supra* note 41 and accompanying text.

382. Treas. Reg. §1.451-2(a) (2005).

383. Furstenberg v. Comm'r, 83 T.C. 755, 791 (1984).

384. The owner of an IRA who has not yet attained age 59-1/2 can withdraw amounts to pay her child's or grandchild's "qualified higher education expenses" without triggering the 10% premature withdrawal tax of section 72(t) of the I.R.C.

385. The IRS has concluded that a Payment Trigger based on changing from full-time to part-time would not immediately cause the NQDC benefits to be taxable. Rev. Rul. 60-31, 1960-1 C.B. 175, 178.

386. I.R.C. § 409A(a)(2)(A)(vi) (2005).

387. *See* Kennedy, *supra* note 362, at A-31 to -36; Janine H. Bosley & Martha L. Hutzelman, *Qualified Plans—Taxation of Distributions*, *in* 370-3d TAX MANAGEMENT PORTFOLIOS, at A-79 to -81 (BNA 2003).

388. I.R.C. § 409A(a)(4)(C) (emphasis added).

389. *Id.* § 409A(a)(4)(C)(ii)-(iii).

A "subsequent election" occurs when a NQDC benefit is further delayed. For example, if the participant defers an amount of compensation and the NQDC arrangement provides for payments to begin in Year 10, a "subsequent deferral" occurs if the participant elects to have the payments begin in Year 11. The Tax Section of the American Bar Association acknowledged that it may be appropriate to tax the participant if she has the right to make this election shortly before the benefit would otherwise become payable (i.e., less than six months).[390] Section 409A not only adopts a lengthier advance period (one year), but also forces the first payment to be extended for at least an additional five years. This "One Year/Five Year Rule" will create significant problems for participants who may actually need the money some day, but will impose no burden on wealthy executives who will not need the deferred money for retirement.

EXAMPLE #15. Middle Manager views her NQDC as an integral part of her retirement plan, along with her Social Security benefits and 401(k) money. During her prime working years, she designates that her NQDC should be paid over a relatively short period of time (for example, from age 65 to 70). She may adopt this strategy because she must begin receiving minimum distributions under her 401(k) plan at age 70-1/2.[391] At 64, Middle Manager reviews her situation and concludes that receiving the NQDC over her life expectancy would be more prudent.[392] However, under Section 409A(a)(4)(C), her plan to be more fiscally responsible will be frustrated. The new Section 409A rule will force Middle Manager to choose between (i) staying with the original distribution plan selected, or (ii) receiving nothing under her NQDC plan for five years. If Middle Manager needs some of the NQDC money in that five-year period, she is stuck with her original choice.

EXAMPLE #16. Fat Cat chose the same NQDC distribution pattern as Middle Manager during his most active working years, but Fat Cat will not need any of the NQDC money for at least six years. Section 409A(a)(4)(C) poses no problem for Fat Cat because he is wealthier than Middle Manager.

The One Year/Five Year Rule is more difficult to avoid, and even more troubling from a tax policy perspective, primarily because Section 409A eliminates the ability to extend the payment schedule by either mutual agreement of the participant and the employer, or unilateral action of the employer.[393] Prior to Section 409A, if amounts would soon become payable,

---

390. *ABA Report, supra* note 127, at 235 ("[I]f the subsequent election relates to the time for payments to commence, a minimum advance period of six months would be required . . . .").

391. I.R.C. § 401(a)(9)(C)(i)(I).

392. A participant may desire a subsequent deferral any time her economic situation is not what she had anticipated when she first deferred the money. For example, in Year 1 an employee may believe she will need the money in a lump sum at retirement. In Year 9 she may realize that it would be preferable to spread out the payments over her life expectancy.

393. Section 409A allows only six Payment Triggers, with the key Payment Trigger being

the parties could agree to extend the payment date.[394] By changing this rule, the One Year/Five Year Rule further reflects Section 409A's disregard of the fundamental principles of the constructive receipt doctrine, which were carefully developed over more than fifty years.[395] Because the constructive receipt doctrine applies the "ability to pay" rationale to a cash basis taxpayer, a taxpayer should be taxed on a benefit when she has an unrestricted right to withdraw the cash or property by giving notice. At that point the right to receive the amount is as good as cash.[396] The Tax Court has recognized that a taxpayer's right to receive a benefit is not as good as cash today if she must obtain the consent of a potentially adverse party to receive it today.[397]

Moreover, this One Year/Five Year Rule will fail to treat the parties to a compensation arrangement who defer payments similarly to the parties involved in a sale of property who defer payments.[398] A situation in which an employer and employee agree to further extend the scheduled payment dates under a NQDC arrangement is analogous to when the buyer initially delivers a promissory note as consideration for the purchase of a capital asset and the parties subsequently agree to extend one or more payment dates under the promissory note. In the case of a sale of property, the parties have substantial flexibility in extending the maturity date without triggering a taxable gain.[399]

Another troubling aspect of the One Year/Five Year Rule is that the Triple Tax may apply when the employer cannot make scheduled payments for valid business reasons—such as cash flow difficulties— but does not delay the first payment for at least five years. For example, if payments are due in Year 9 and the employer does not make the payment because of cash flow problems, but the parties agree that the payments can be made in Year 10, it appears that a Section 409A violation has occurred. It is inappropriate to impose the Triple Tax on the employee because the employer failed to honor its obligation.

"a specified time (or pursuant to a fixed schedule) specified under the plan *at the date of the deferral of such compensation.*" I.R.C. § 409A(a)(2)(A)(iv) (emphasis added). If the employer and employee mutually agree to change the payment dates after the amounts have been deferred, the new payment schedule would not be a permitted Payment Trigger and would violate Section 409A and trigger the Triple Tax.

394. *See supra* notes 204, 213 and accompanying text (discussing the *Veit I, Veit II,* and *Oates* cases).

395. *See supra* notes 183-203 and accompanying text.

396. *See supra* notes 183-203 and accompanying text.

397. *See* Metcalfe v. Comm'r, 43 T.C.M. (CCH) 1393, 1398-99 (1982).

398. The concept that taxpayers in similar situations should be treated the same is described as "horizontal equity." *See* NEWMAN, *supra* note 41, at 26.

399. Treas. Reg. § 1.1001-3(e)(3)(ii) (2005) ("The deferral of one or more scheduled payments within the safe-harbor period is not . . . material [and will not trigger a taxable event.] The safe-harbor period . . . extends for a period equal to the lesser of five years or 50 percent of the original term . . . .").

E. An Employee May Pay a 20% Additional Tax on an Amount She Never Receives, and May Be Unable to Receive a Refund

An employee may pay the 20% additional tax on a promised NQDC benefit, and the employer may never pay the benefit because, for example, the employer declares bankruptcy. A mechanism should be added to permit the employee to obtain a refund of the 20% additional tax.

VI. POLICY THEMES THAT SHOULD BE CONSIDERED IN FUTURE ATTEMPTS AT COMPENSATION REFORM

Because Section 409A merely creates timing rules while the fundamental compensation abuses and their causes remain, this Section discusses policy themes that should be considered during future attempts at compensation reform.

A. Failure of Boards of Directors and Compensation Committees at Publicly-Traded Corporations to Negotiate at Arm's-Length Over Executive Compensation

A typical debate over whether compensation paid to a service provider is grossly excessive might go something like this:

> "It's crazy that the Miami Heat is willing to pay Shaquille O'Neal $27.7 million in 2005 to play basketball."[400]
>
> "They would not pay it if he wasn't worth it."
>
> "But not even a brain surgeon makes that much money."
>
> "Yes, but the owners of the Miami Heat are not dumb. Every dollar they pay Shaquille is an expense that reduces their profit, so if they could get him for less, they would have never offered that much."

This debate simply acknowledges the common sense notion that people do not waste their own money in transactions with strangers. The IRS and the courts have consistently recognized this principle in deciding "reasonable compensation" cases.[401] If a closely-held corporation pays compensation to an

---

400. *See* Bob Sansevere, *NBA Needs to Learn from NHL*, ST. PAUL PIONEER PRESS, May 31, 2005, at C1 ("That works out to $73,972.60 a day, every day of the year, in season and out. In the time it takes Shaq to sneeze, he has made about $2.57.").

401. Section 162(a)(1) permits an employer to claim an income tax deduction for reasonable compensation paid in the ordinary course of a trade or business. If a compensation arrangement involves an employee with family ties to the business, the IRS and the courts are much more likely to scrutinize the transaction because the employer may be paying excess compensation disguised as a gift. For example, if a mother owns all the stock of a corporation,

executive who has no family relationship to the owners, the IRS and the courts are unlikely to challenge the reasonableness of the compensation payment, because an extra $1 paid to the unrelated executive will reduce the corporation's (and the shareholders') profits by $1.

1. Reasons Why Directors Do Not Bargain at Arm's-Length

The natural tension that is normally present when employees and employers negotiate compensation is absent in the case of top executives at publicly-traded companies. The *shareholders* have a direct financial interest in the corporate profits; consequently, they have a strong motive to pay as little compensation as possible to the top corporate executives. But the shareholders do not directly negotiate with the top executives over compensation. Instead, that job falls to the corporation's Board of Directors or the Compensation Committee or both. Typically the directors or committee members own a minuscule portion of the publicly-traded corporation's stock.

> [D]irectors commonly bear only a negligible fraction of the cost imposed by flawed compensation arrangements. Consider, for example, a director who owns 0.005 percent of the company's shares. And suppose that the director is contemplating whether to approve a compensation arrangement requested by the CEO that would reduce shareholder value by $10 million. Given the director's fraction of total shares, the reduction in the value of the director's holdings that would result from approval of the CEO's request would be only $500. Such a cost or even one several times larger, is highly unlikely to overcome the various factors exerting pressure on the director to support the CEO's request.[402]

The Board of Directors and the Compensation Committee are negotiating with "other people's money." Commentators have identified several economic and social-psychological factors that lead members of the Board of Directors and Compensation Committees to award excessive compensation to top executives, as discussed below.

a. Business with the Corporation

The CEO has substantial control over the corporation's resources[403] and may cause the corporation to do business with an entity owned by the director.

---

and the corporation pays compensation to the daughter, the IRS and the courts may scrutinize the transaction to determine if the mother is using the compensation to make indirect gifts to the daughter. *See* Kafka, *supra* note 54, at A-12 ("Virtually all challenges by the IRS to the deductibility of compensation have occurred in the context of salary arrangements between related parties, involving dealings . . . between corporations and shareholders or relatives of shareholders . . . .").

402. BEBCHUK & FRIED, *supra* note 17, at 34.

403. *Id.* at 27.

For example, "Verizon's 2001 board included an executive director of Boston Consulting Group, which received $3.5 million from Verizon for services in 2000; the CEO of a railroad that was paid $650,000 by Verizon for services and products; and two attorneys from law firms that provided Verizon with legal services."[404]

While some directors may not do business with the corporation during their tenure as a director, either because of a conflict of interest policy or a desire to maintain some appearance of independence, those directors may hope to do business with the corporation at the end of their tenure.[405] A director who hopes to gain business from the corporation now or in the future has a natural inclination to avoid a fight with the CEO over compensation. "This . . . highlights how difficult it is to prevent directors who have ties to a business from being influenced by the CEO's power over the company's resources."[406]

b. The Lack of Information and the Role of Compensation Consultants

Corporations frequently hire compensation consultants to provide compensation surveys and other information to the corporation's board of directors or to assist the compensation committee in evaluating compensation proposals. Since directors and compensation committees may lack time, information, and expertise, they may rely heavily on advice from outside consulting firms when making compensation decisions.[407] Even if an independent committee hires the compensation consultant, the bulk of the consultant's work typically will come from the corporation's human resources department, which is under the CEO's control.[408] Thus, the compensation consultant will have an economic incentive to produce information that supports high compensation for top executives.[409] Further, because its income is not linked to shareholder value, the compensation consultant will not bear the cost of top management being overcompensated. As a result, "consultants [can] only benefit from using their discretion to favor the CEO."[410] Warren Buffett once stated that compensation consultants "[have] no trouble perceiving who buttered their bread."[411]

---

404. *Id.* at 27-28.

405. *See id.* at 29 ("[C]onsider a member of the company's board who is a partner in a law firm that is not currently providing services to the company. This lawyer still has an economic incentive to remain on good terms with the CEO because the company could be a future client. . . . [T]he possibility of such future business might affect the lawyer's economic incentives even while the lawyer is still an independent director.").

406. *Id.*

407. *Id.* at 38.

408. *Id.*

409. *Id.*

410. *Id.*

411. *See id.* at 38.

c. Friendship and Loyalty Considerations

"Many independent directors have some prior social connection to, or are even friends with, the CEO or other senior executives."[412] But even an independent director with no prior connection to the CEO may be grateful to a CEO who has helped bring him onto the board, which may include simply "not blocking the director's nomination."[413] "With such a background, directors often start serving with a reservoir of good will toward the CEO, which will contribute to a tendency not to bargain aggressively with the CEO over pay."[414]

d. Collegiality and Team Spirit

The CEO "is generally a fellow member of the board and often its chair."[415] The social rules of boards "tend to foster board cohesion."[416]

> Perhaps once or twice a year, members of the compensation committee must take off their hats as colleagues of the CEO and put on their hats as arm's-length bargainers with the CEO over his or her compensation. This switch is likely to be difficult even for well-meaning directors attempting to represent shareholders' interests in these negotiations. Evidence indicates that individuals working with a group feel pressure to placate group members, often at the expense of interests not directly represented at the table. There is no reason to believe that the members of the compensation committees are immune to such pressure.[417]

e. Cognitive Dissonance

Approximately two-thirds of the directors of public companies who are members of such corporations' compensation committees are or were top executives of other corporations.[418]

> The compensation decisions of directors who are or were executives themselves are likely to be affected by cognitive dissonance. Individuals are known to develop beliefs that support positions consistent with their self-interest. These beliefs enable individuals to avoid the discomfort of enjoying benefits that they believe to be undeserved.
> An executive or former executive who has benefited from generous and favorable pay arrangements is thus likely to have formed a belief that such arrangements are desirable and serve shareholders.[419]

---

412. *Id.* at 31.
413. *Id.*
414. *Id.*
415. *Id.* at 32.
416. *Id.*
417. *Id.* (footnote omitted).
418. *Id.* at 33.

f. CEO's Ability to Influence the Nomination of Director

In light of the actual and potential financial benefits that flow from being a director, incumbent (and "want-to-be") directors have a financial interest in pleasing those who control the process of electing the directors. While shareholders technically can vote to reject a slate of directors, "[d]issident shareholders contemplating putting forward their own director slate have confronted substantial obstacles. As a result, the director slate proposed by the company['s nominating committee] has almost always been the only one on the ballot [submitted to shareholders.]"[420] Even when CEOs (or their subordinates) do not directly serve on nominating committees,

> Boards and nominating committees have been unlikely to nominate a director clearly opposed by the CEO. *At a minimum, CEOs have had considerable power to block nominations.* Thus, sparring with the CEO over executive compensation could have only hurt a director's chances of being renominated to the board. "Going along" with the CEO's pay arrangement has been a much safer strategy.
>
> . . . .
>
> [Even if the nominating committee is composed of independent directors, as required by NYSE listing requirements], the CEO's wishes can be expected to continue to influence the decisions of the nominating committee; after all, the directors appointed to the board will have to work with the CEO. Indeed, experts interviewed by the *Wall Street Journal* have advised boards to consult with management on the independent nominating committee's choices. And, as a lawyer who has served on the board of several public companies said, "I think as a practical matter, few new directors would accept without knowing that the CEO is enthusiastic about the decision. . . . No one likes to go to the boardroom thinking they've been imposed on the CEO."[421]

Dick Grasso, former CEO of the NYSE, allegedly "personally selected which board members served on the compensation committee, and some directors he selected were those with whom he had friendships or personal relationships."[422]

---

419. *Id.* (footnote omitted).

420. *Id.* at 25 (footnote omitted).

421. *Id.* at 26 (quoting Daniel Nasaw, *Opening the Board: The Fight is on to Determine who will Guide the Selection of Directors in the Future*, WALL ST. J., Oct. 27, 2003, at R8) (emphasis added); *see also* NYSE, Inc., Listed Company Manual § 303A.04 (2005) (setting forth the rule that nominating committees must be composed entirely of independent directors).

422. Daniel Dunaief, *Grasso Paid Drivers, Assistant Big Bucks*, DAILY NEWS (New York), Feb. 3, 2005, at 66.

g. Director Compensation

The compensation paid by a corporation to its directors can be substantial, and in recent years, their compensation rates have risen. "In 2002, director compensation averaged $152,000 in the largest 200 companies and $116,000 in the largest 1,000 companies."[423] Since 2002, director compensation has continued to rise; in 2005, it is expected to reach an average of $200,000 per director of the Fortune 200.[424] The typical director works approximately 200 hours per year,[425] resulting in an annual salary that equals approximately $508 per hour. "As the Company leader, usually as a board member, and often as board chair, the CEO has some say over director compensation. . . . Independent directors who are generous with the CEO might reasonably expect the CEO to use his or her bully pulpit to support higher director compensation."[426] Both dependent and independent directors may also expect the CEO's support during the selection of Board members.

h. Director Perks

In addition to cash or stock compensation, directors may receive other benefits from the corporation. For example, "directors of UAL Corp. (which owns United Airlines) can fly United free of charge, and directors of Starwood Hotels get complimentary nights in company hotels."[427] "WorldCom CEO Bernard Ebbers allowed . . . the director who chaired the [WorldCom] compensation committee, to rent a Falcon 20 jet for $1 a month plus $400 an hour [and] minor expenses . . . . The value of the lease was between $1.4 million and $3.4 million . . . ."[428]

---

423. BEBCHUK & FRIED, *supra* note 17, at 25.

424. Gary Strauss, *Board Pay Gets Fatter as Job Gets Hairier*, USA TODAY, Mar. 7, 2005, at 1B ("If early patterns hold, the average director's pay at Fortune 200 companies could surge to $200,000, up 14% from 2004 . . . ."). Director compensation can include a fixed cash amount ($175,000 per year at Alcoa), meeting fees ($2,000 per meeting at utility Pepco), annual stock grants (4,000 shares at Morgan Stanley) and extra cash fees for serving as chair of a committee ($20,000 for serving as chair of the audit committee at U.S. Bancorp). *Id.* In summary, "[D]irectors are well compensated." *Id.*

425. BEBCHUK & FRIED, *supra* note 17, at 37 ("[I]n 2002 directors at Fortune 1,000 companies spent about 190 hours on board service . . . .") (citing Judith Burns, *Everything You Wanted to Know About Corporate Governance . . . But Didn't Know to Ask*," WALL ST. J., Oct. 27, 2003, at R6)).

426. *Id.* at 30.

427. *Id.* at 25.

428. *Id.* at 221 n.18 (citing Susan Pulliam et al., *WorldCom Board Will Consider Rescinding Ebbers's Severance*, WALL ST. J., Sept. 10, 2002, at A1; Christopher Stern, *WorldCom Director Urged to Leave*, THE WASHINGTON POST, Sept. 13, 2002, at E01).

i. Business Connections

Membership on the Board can provide a director with prestige, as well as business opportunities through exposure to other decision makers.[429]

j. Charitable Contributions

"CEOs have used their power to direct [charitable] contributions to benefit directors."[430]

2. The Exchange Listing Rules Will Promote Independence, But Likely Will Not Inspire Arm's-Length Bargaining

The New York Stock Exchange and other exchanges have amended their listing requirements to increase director independence:[431] "Under the NYSE rules, a majority of all directors must be 'independent,' as must all directors on audit, *executive-compensation*, nominating and corporate-governance committees."[432] Although only time will tell, these rules likely will be insufficient to offset the factors that discourage directors and compensation committees from engaging in arm's-length negotiations over executive compensation.[433] While the new rules require that listed companies have a

---

429. *Id.* at 25.

430. *Id.* at 28 (discussing contributions by Verizon to the National Urban League, whose head sat on Verizon's board, a contribution by Oracle to Stanford while three Stanford University professors were on the Board, and donations Enron made to some of its directors' favorite charities).

431. *See* NYSE, Inc., Listed Company Manual § 303A (2005); NASD Rule 4350 (2005); American Stock Exchange Company Guide sec. 121, 801-09; Exchange Act Release No. 3448785 (Nov. 3, 2004); Exchange Act Release No. 34-48863 (Dec. 1, 2003).

432. John R. Emshwiller & Joann S. Lublin, *In Boardrooms, 'Independent' Is Debatable*, WALL ST. J., Mar. 3, 2005, at C1 [hereinafter Emshwiller 1]. The NYSE's rules require that all directors on the nominating committee be independent, while "NASDAQ's new listing provisions require that director nominees be selected or recommended either by a majority of the independent directors or by a nominating committee composed solely of [independent] directors." BEBCHUK & FRIED, *supra* note 17, at 26.

433. BEBCHUK & FRIED, *supra* note 17, at 23 ("[These stock exchange requirements] are unlikely to eliminate substantial and widespread deviations from arm's-length contracting."); *id.* at 25 ("[T]here are good reasons to doubt that the mere presence of independent directors . . . can ensure a pay-setting process that approximates arm's-length bargaining."). Even before the stock exchange rules were enacted in 2004, SEC rules required public companies to report annually instances in which a director, senior executive or close relatives of either receives more than $60,000 from the company. John R. Emshwiller & Joann S. Lublin, *Citigroup Discloses Relatives of Executives Were on Payroll*, WALL ST. J., Mar. 3, 2005, at C4 [hereinafter Emshwiller 2]. After Walt Disney Co. was cited for violations of this rule, Linda Chatman Thomas, SEC deputy enforcement director, stated, "Failure to comply with the SEC's disclosure rules in this area impedes shareholders' ability to evaluate the objectivity and independence of

compensation committee consisting solely of independent directors,[434] the "$1 Million Cap" of Section 162(m) of the I.R.C., which has been in effect since 1994, has effectively required that publicly-traded companies use a compensation committee consisting solely of independent directors.[435] The $1 Million Cap rule provides that a publicly-traded corporation can only deduct compensation in excess of $1 million paid to one of its five highest paid executives[436] if such compensation is based on the achievement of performance goals,[437] and the performance goals are established by a *compensation committee consisting solely of independent directors*.[438]

Also, the exclusions and exceptions of the new listing rules[439] highlight the ability of top executives to influence even independent directors. The following directors are considered "independent" under these rules:

CORPORATION	EXECUTIVES' FINANCIAL CONTROL OVER AN "INDEPENDENT" DIRECTOR
BB&T Corp (a financial services holding company)	"[A]n attorney whose law firm works for the [company] heads the board committee that sets executive pay . . . ."[440]
Coca-Cola	Warren Buffet is listed as an "independent" director even though he is chairman and CEO of Berkshire Hathaway, Inc., which did more than $100 million of business with Coke in 2003.[441]

directors." *Id.*

434. *See* Emshwiller 1, *supra* note 432, at C1.

435. I.R.C. § 162(m) (2005).

436. Treas. Reg. § 1.162-27(c)(2) (2005).

437. I.R.C. § 162(m)(4)(B)-(C).

438. *See* Treas. Reg. § 1.162-27(e)(3) (2005) (describing the requirements for being "independent").

439. Key exceptions include directors who are executives at entities that do business with the listed company when the total amount paid is less than the *greater* of (i) $1 million, or (ii) 2% of the entity's revenue. NYSE, Inc., Listed Company Manual § 303A.02(b)(v) (2005).

440. Emshwiller 1, *supra* note 439, at C1. A spokesperson for BB&T Corporation argued that the fees paid by BB&T to the director's law firm were "immaterial to his firm." *Id.* In a review of the corporate filings of 150 listed companies, the Wall Street Journal found 20 instances in which an attorney was (i) on the corporation's board of directors; (ii) worked for a law firm that is employed by the company; and (iii) is classified as "independent" under the exchange rules. *Id.*

441. *Id.* In response to a Wall Street Journal inquiry Mr. Buffet declined to comment, but it has been observed that the $100 million from Coca-Cola was "well under 2% of Berkshire's 2003 annual revenue of more than $60 billion." *Id.*

General Motors	The CEO of Merrill Lynch & Co. is considered an "independent" director because "GM's underwriting and investment banking business provides well under 2% of Merrill's annual revenue."[442]
Citicorp	Two directors are "independent" under the rules even though they have children working for Citicorp who receive $350,000 and $236,500, respectively, in annual salary.[443]

Also, commentators note,

> [The NYSE] rules do not prohibit a firm from giving directors compensation on top of [regular director's] fees; rather, they only limit the amount of such compensation . . . to $100,000 a year in such additional compensation, hardly a negligible amount for many directors. Moreover, . . . compensation given to an immediate family member who is a nonexecutive employee of the company would not count toward this $100,000 limit.[444]

Furthermore, even a director who is independent today may desire to obtain work from the corporation in the future; bickering with the top executives over compensation likely will not improve a current director's chances of gaining the company's business at a future date.[445]

B. Excessive Contingent Compensation, Including Stock Options

The $1 Million Cap, which imposes certain limitations on executive compensation, has two glaring weaknesses. First, the penalty for violating the $1 Million Cap falls on the corporation (and indirectly on the shareholders), and not on the overpaid executive—the corporation is denied an income tax deduction for excessive compensation.[446] Second, the $1 Million Cap does not apply to "performance-based compensation" if certain requirements are met.[447]

As a result, the compensation package of a top executive is often structured in a bizarre manner to avoid violating the $1 Million Cap. While total annual

---

442. *Id.*

443. *Id.*; Emshwiller 2, *supra* note 433, at C4.

444. BEBCHUK & FRIED, *supra* note 17, at 28 (citing NYSE, Inc., Listed Company Manual § 303A.02(b)(ii) (2005)).

445. BEBCHUK & FRIED, *supra* note 17, at 29; *see also supra* notes 440-43 and accompanying text (including the example of an attorney serving on the board of directors).

446. I.R.C. § 162(m)(1) (2005). In contrast, when an officer of a public charity is overpaid, an excise tax will be imposed on the officer. *See id.* § 4958(a)(1).

447. *Id.* § 162(m)(4)(B)-(C).

compensation of CEOs of S&P 500 firms averages $14.7 million,[448] their *salary* averages just more than $2 million.[449] Since the $1 Million Cap only permits a corporation to deduct $1 million of fixed salary paid to a top executive, but places no limit on the deductibility of performance-based compensation,[450] such compensation (i.e., bonuses and stock options) constitutes the lion's share of executive compensation. Ridiculously large amounts of compensation can be paid under the guise of "performance-based" compensation, with the amount of compensation having little relationship to increases in shareholder value.[451]

1. Bonuses

While "[t]he term 'bonus' suggests a payment for particularly good performance,"[452] bonuses are often granted for mediocre or even poor performance. A survey found that at 100 major U.S. corporations, the average CEO received a bonus that exceeded 140% of her salary.[453] Bonuses are often granted based on a comparison of performance to budget or on a comparison to the prior year's performance.[454] A bonus scheme based on prior year's performance can lead to especially peculiar incentives. An especially good performance this year will make it more difficult to obtain a bonus next year, while a poor performance this year will make it easier to obtain a bonus next year.[455] "[A] large majority of companies with bonus plans based on objective measures do not base bonuses on the firm's performance relative to its peer group."[456]

When investments are favorable, some corporations include the earnings on pension funds when calculating profits for purposes of determining the executive's bonus, even though the investment return on pension assets has nothing to do with the executive's performance.[457] Not surprisingly, when

448. BEBCHUK & FRIED, *supra* note 17, at 1.

449. *Id.* at 122.

450. *See supra* text accompanying notes 446-47 (discussing Section 162(m)). Performance-based compensation must be "reasonable" to be deductible, I.R.C. § 162(a), but the IRS and the courts have struggled in applying the "reasonable compensation" standard. *See* Kafka, *supra* note 54, at A-3 ("[E]very reasonable compensation inquiry is resolved by reference to the particular facts and circumstances of a given case.").

451. BEBCHUK & FRIED, *supra* note 17, at 124-25.

452. *Id.* at 124.

453. Lublin, *supra* note 33, at A1 ("[But] [c]lerical and technical-support staff earned an average bonus of 5% of salary [in 2004] at concerns granting bonuses across the board . . . .").

454. BEBCHUK & FRIED, *supra* note 17, at 124.

455. *Id.*

456. *Id.*

457. A particularly egregious situation has been described as follows:

> In 2001, Verizon Communications reported a net income of $389 million and awarded its executives bonuses based on that amount. Net income would have been negative, however, had the company not included $1.8 billion of pension income. Thus, Verizon

pension funds lose money, the bonus calculation can be revised to exclude pension investments from the bonus calculation.[458]

> [A]lthough most bonus payments are regarded as performance based for purposes of [the "$1 Million Cap"], they often are only weakly tied to performance. . . . [B]onuses are often conditioned on easily attained performance targets that do not reflect good performance relative to peer firms. They also often reward executives for things—positive developments in the market or the sector as a whole, other types of luck, making acquisitions, and so forth—that clearly have nothing to do with managerial performance.[459]

Bonuses appear to be rising—a survey of 100 major U.S. corporations concluded that CEO bonuses rose over 46% in 2004.[460] The Walt Disney Company's Board of Directors granted Michael Eisner a $7.25 million cash bonus in the same year that Disney's shareholders "revolt[ed]" and removed Mr. Eisner as chairman of the Board at Disney.[461]

2. Stock Options

In contrast to bonus formulas that generally must be designed and manipulated to provide excessive compensation, stock options, by their very nature, tend to award executives regardless of their performance. A stock option is simply a right to buy the employer's stock at a fixed price (usually the current market price at the time the option is granted) for a fixed period of

---

was able to use pension earnings to convert [a loss to a profit], giving the firm cover to provide managers with higher bonuses.

It gets worse. It turns out that Verizon's pension funds did not generate any real income in 2001; they had negative investment returns, losing $3.1 billion in value. How, then, could Verizon report income of $1.8 billion from its pension assets? The company merely increased its projection of future returns on pension assets to 9.25 percent, a move allowed under the accounting rules then in effect. Thus, the $1.8 billion in pension income used to move Verizon into the black did not even reflect actual returns generated by the pension funds. The pension income was simply the result of a change in the accounting assumptions. This certainly did not create any value for the firm or its shareholders.

BEBCHUK & FRIED, *supra* note 17, at 125; *see also* Floyd Norris, *Pension Folly: How Losses Become Profits*, N.Y. TIMES, Apr. 26, 2002, at C1 (discussing Verizon's alleged pension plan profits). Other firms that have used earnings on pension funds to boost executive bonuses include General Electric and IBM. *Id.*

458. *See* Jesse Drucker & Theo Francis, *Pensions Fall—Not CEO's Bonus*, WALL ST. J., June 18, 2003, at C1 ("Now that the up-and-down stock market has made many pension plans a drag, not a boost, on earnings, some companies including General Electric Co., Delta Air Lines and Verizon Communications Inc. have started removing pension effects from their executive compensation formulas.").

459. BEBCHUK & FRIED, *supra* note 17, at 135.

460. Lublin, *supra* note 33, at A1.

461. *Id.*

time.[462] "Warren Buffett has said: 'There is no question in my mind that mediocre CEOs are getting incredibly overpaid. And the way it's being done is through stock options.'"[463] Almost 75% of the total compensation paid to the Enron Top 200 was in the form of stock options—a total of over $1.06 *billion*, or an average of $5,317,500 to each of the Enron Top 200 in the year immediately before Enron went bankrupt.[464]

A major problem with using stock options as compensation is that the CEO may have performed poorly (as compared to other CEOs), and the corporation may have performed badly within its industry, yet the CEO may still receive a windfall.[465]

> [C]hanges in share price are not a good indicator of a manager's own performance. A company's stock price can increase for reasons that have nothing to do with its managers' own efforts and decision making. Falling interest rates, for example, can cause stock prices to increase considerably without managers lifting a finger. Indeed, one study of U.S. stock prices over a recent ten-year period reported that only 30 percent of share price movement reflects corporate performance; the remaining 70 percent is driven by general market conditions. If performance is measured by changes in share price, managers who perform poorly relative to their peers might still be rewarded when the market or sector rises as a whole.[466]

Although eliminating all factors unrelated to management's performance that impact share price would be difficult, "identifying those due to sector or general stock market trends is a more straightforward exercise."[467] Basically, a company could simply increase the price that the executive must pay to exercise

---

462. EXAMPLE #17. Big Co., a publicly traded company, grants its CEO, Ernie Executive, an option to purchase 1,000 shares at its current market price, $10 per share, for the next 10 years. On the date of grant, Ernie has no incentive to exercise the option; if he exercises the option he will need to pay $10,000 to acquire 1,000 shares of Big Co. There is no need for Ernie to exercise the option because even without exercising the option, Ernie can acquire 1,000 shares for $10,000. In contrast, in year 9, the market value of Big Co. stock is $19 per share and Ernie has an incentive to exercise the option. In exercising the option, Ernie will pay $10,000 for 1,000 shares, even though they are now worth $19,000. In effect, Ernie received an extra $9,000 in compensation because of the option.

463. BEBCHUK & FRIED, *supra* note 17, at 143.

464. *See supra* notes 90 and 94 and accompanying text (this figure likely represents the value of options exercised in 2000).

465. BEBCHUK & FRIED, *supra* note 17, at 143.

> [D]uring the big market boom of the 1990s, even executives with subpar performance often reaped large gains from options. With conventional options, if the market goes up 300 percent, an executive whose firm lags the market by 50 percent will still make very large profits—larger profits, indeed, than an executive whose firm beats the market by 50 percent during a period when the market is flat.

Id.

466. *Id.* at 139 (footnotes omitted).

467. *Id.* at 140.

the option to purchase the stock (called the "exercise price") by the overall percentage increase of all stocks sold on the exchange (or the overall percentage increase of all stocks sold on the exchange in the same industry). This technique is called "indexing the exercise price,"[468] and although it is easy to implement,[469] it is seldom used.[470]

C. *Recognizing Distinctions Between Closely-Held Corporations and Publicly-Traded Corporations*

A key step in a legislative or regulatory effort should be to identify the appropriate behavior to be regulated. If a statute or regulation is needed to discourage certain behavior, then the statute or regulation should focus on that particular behavior. In developing Section 409A, the legislature appears to have ignored this fundamental principle. The Enron Compensation Report, which precipitated the enactment of Section 409A, was motivated by valid concerns over egregious compensation practices at the seventh largest corporation in the United States. In particular, the Enron Compensation Report revealed that the Enron Top 200 each received over $5 million in allegedly performance-based compensation, in the form of stock options, in the year immediately before the company went bankrupt.[471] The report also revealed that the board of directors, the compensation committee, and the outside compensation consultants all "rubber-stamped" the compensation proposals from Enron's management.[472]

Amazingly, the Enron Compensation Report made no recommendations regarding stock options[473] and proposed no changes to the structure of compensation approval at publicly-traded companies. Ultimately, the Enron Compensation Report led to enactment of Section 409A, which applies equally to publicly-traded and closely-held corporations.[474]

The compensation process at closely-held corporations is very different from the compensation process utilized by publicly-traded corporations. At

---

468. *Id.* at 141 ("The most familiar way to reduce windfalls is by indexing the exercise price—having it rise and fall either with sector or broader market movements, thereby screening out the effects of those movements on a firm's stock price.").

469. *Id.* at 141-43.

470. "In 2002, only 8.5 percent of large public firms issuing options to executives conditioned even a portion of the grant on performance." BEBCHUK & FRIED, *supra* note 17, at 143.

471. *See supra* note 90.

472. *See supra* notes 103-09 and accompanying text.

473. Enron Compensation Report, *supra* note 2, at 41 ("In implementing its stock-based compensation programs, Enron appeared generally to follow IRS published guidance. Thus, no recommendations are made with respect to such programs.").

474. *See supra* note 12. The one distinction between closely-held and publicly-traded corporations under Section 409A is that upon a "separation from service," distributions cannot be made to top executives at publicly-traded corporations for 6 months. *See* I.R.C. § 409A(a)(2)(B)(I) (2005).

closely-held corporations, the shareholders and their families may directly serve on and control the nominating committee, the compensation committee, and the Board of Directors. Thus, the directors at a closely-held corporation typically have an affinity of interest with the shareholders—they have a direct economic incentive to keep executive compensation as low as possible. But at publicly-traded companies, the directors often will own a tiny fraction of the corporation's stock, and when they negotiate executive compensation, they are spending other people's money. Unfortunately, there is no meaningful distinction between how Section 409A applies to closely-held corporations and publicly-traded corporations. In the future, if Congress attempts meaningful compensation reform, distinctions between closely-held corporations and publicly-traded corporations must be considered.

D. Increasing the Progressivity of Income Tax Rates

While progressivity is a much broader issue than compensation reform, it exacerbates the unfairness of excessive compensation. In effect, top executives are overpaid and they do not pay their fair share of taxes. The "ability to pay" principle[475] is at the heart of the argument for greater progressivity. If a certain amount must be raised in taxes, a greater proportion should be extracted from those best able to afford it. In the past, the federal income tax rates have been much more progressive. During World War II, the government enacted the "Victory Tax," and at one time, the maximum federal income tax rate was 94%.[476] Thus, at that time, highly paid corporate executives would have paid 94% of their excess compensation in taxes.

The Tax Reform Act of 1986 greatly reduced progressivity and changed the top federal income tax rate from 50% to 28%.[477] Proponents of the shift to a "flatter" tax system argued that the reduction in the top marginal rate would greatly reduce tax avoidance and evasion by the wealthy.[478] On the contrary, it can be argued that the 1990s saw the greatest proliferation of tax shelters in history,[479] and the flat tax movement has simply resulted in the rich paying less

475. *See supra* note 41 and accompanying text.

476. *See* NEWMAN, *supra* note 41, at 11 ("The big change in taxation came as a result of World War II. The government needed money, so they raised the rates, at one point to a high marginal rate of 94%.").

477. Pub. L. No. 99-514, § 101(a), 100 Stat. 2096 (1986). Progressivity has increased gradually since 1986, with the top rate moving to 39.6%. *See* I.R.C. § 101(a)-(d).

478. *See* Roscoe L. Egger, *Introduction* to THE PRICE WATERHOUSE GUIDE TO THE NEW TAX LAW ix (1986) ("The new tax code will combine significantly lower rates for individuals and businesses, with a tax base that is expanded through the repeal or modification of many of the incentives initially designed to encourage social or economic goals. This code will treat all taxpayers more fairly: Under the new rules, it will be virtually impossible for any profitable company or wealthy individual to avoid paying tax.").

479. *See e.g.*, Jonathan D. Glater, *Settlement Seen on Tax Shelters by Audit Firm*, N.Y. TIMES, Aug. 27, 2005, at A1 ("The settlement is the latest step in the government's investigation of questionable shelters, which were created and sold by accounting firms in the late 1990's,

tax and getting richer. While the absence of meaningful compensation reform inflates the income of top executives, the lack of progressivity allows the overpaid to retain an unfair portion of their excessive compensation.

VII. CONCLUSION

The investigation into Enron's compensation practices was warranted based on just a few statistics:

- The total compensation of the Enron Top 200 doubled from 1998 to 1999, and tripled from 1999 to 2000,[480] fueled largely by so-called "performance-based compensation," even though Enron went bankrupt in 2001.
- Compensation based on stock options nearly quadrupled from 1998 to 1999, and more than quadrupled from 1999 to 2000.[481]
- The Enron Top 200 averaged over $5 million each in stock option compensation for the year 2000 alone.[482]
- Ken Lay deferred $32 million in compensation the year before Enron went bankrupt.[483]

The Enron Compensation Report disclosed underlying causes for these compensation abuses—in particular, a complete breakdown in the oversight of executive compensation by Enron's Board of Directors and Compensation Committee, who merely "rubber-stamped" all of management's compensation proposals.[484] The report also emphasized that Towers Perrin and other outside consultants, hired to critically analyze compensation paid to Enron's top executives, supported the compensation proposals from Enron's top management.[485]

Despite the valuable information and analysis in the Enron Compensation Report, Congress' response in Section 409A fails to address the fundamental compensation problems.[486] Although stock option compensation made up almost three-fourths of the total compensation paid to the Enron Top 200 in the

---

and allowed wealthy investors to evade billions of dollars in taxes.").

480. Enron Compensation Report, *supra* note 2, at 545.

481. *Id.* at 547, n.1627.

482. *Id.* ($1,063,500,000/200 = $5,317,500).

483. *Id.* at 604, n.1817.

484. *Id.* at 19.

485. *Id.* at 36 ("[I]n some cases, the [Towers Perrin] studies appeared to be designed to justify whatever compensation arrangement management wanted to adopt.").

486. Income tax rules likely can not provide all the solutions to the fundamental problems. But it appears that Congress and the IRS have expended enormous amounts of time and resources to enact and implement the Section 409A rules, even though NQDC represented less than 5% of total compensation for the Enron Top 200 from 1998 through 2000. *See supra* note 94 and accompanying text.

three years before Enron went bankrupt,[487] the Enron Compensation Report expressly states that the Joint Tax Committee has no recommendations regarding stock-based compensation.[488] Instead, the report focuses on NQDC, which constituted less than 5% of the total compensation to the Enron Top 200 in the three years before bankruptcy.[489] In particular, the report focuses on a perceived abuse regarding NQDC that did not exist. The Report points out that Enron distributed NQDC benefits to its top executives immediately before Enron's bankruptcy, and argues that these distributions damaged the company's other employees, creditors, and shareholders. Unfortunately, the report fails to discuss that the bankruptcy trustee can force these executives to repay all of those NQDC benefits to the Enron bankruptcy estate.[490] While the Enron Compensation Report provides valuable facts for meaningful compensation reform, the statute arising from the report, Section 409A, fails to address the important issues.

If publicity surrounding corporate scandals continues to sustain an environment conducive to meaningful compensation reform, Congress should focus on actual compensation abuses and their causes, such as:

- grossly excessive compensation paid to top executives at publicly-traded corporations, particularly in the form of stock options or other "performance-based" compensation;
- the lack of structures and incentives to require or encourage boards of directors and compensation committees to bargain at arm's-length over executive compensation at publicly-traded corporations; and
- the unfair tax burden imposed by the absence of meaningful progressive income tax rates (for example, a middle-manager with annual taxable income of $180,000 is subject to almost the same marginal income tax rate as a CEO receiving $10 million, $20 million, $30 million or more in compensation each year).

Many challenging issues in the area of executive compensation are worth further pursuit. The SEC chairman has been quoted as saying, "One of the great, as-yet-unsolved problems in the country today is executive compensation and how it is determined."[491] Congress should strive to adopt finely-tuned rules that will curb such compensation abuses without imposing excessive

487. *See supra* note 90 (listing stock option compensation at $1.37 billion out of total compensation of $2.01 billion in the chart on "Compensation Paid to the Top 200 Highly-Compensated Employees for 1998-2000"). Also, stock option compensation in 2000 was $1.06 billion out of total compensation of $1.42 billion. *Id.*

488. Enron Compensation Report, *supra* note 2, at 41.

489. *See supra* note 94 and accompanying text.

490. *See supra* Part II.F.1.

491. BEBCHUK & FRIED, *supra* note 17, at 189 (quoting SEC Chairman William Donaldson in 2003).

transaction costs on others. With regard to Section 409A, the Enron Compensation Report focused on a perceived abuse that did not exist, and Congress wasted valuable time and energy in enacting mere timing rules that will just be an annoyance for publicly-traded companies and their highly-paid executives, while setting a dangerous tax trap for the uninformed at closely-held companies.

SEARCH AND SEIZURE—THE EXPANSION OF POLICE POWERS UNDER THE FOURTH AMENDMENT—THE USE OF FORCE AND THE PROPRIETY OF POLICE QUESTIONING DURING A SEARCH.

Muehler v. Mena, 544 U.S. 93 (2005).

I. INTRODUCTION

Iris Mena and her father, Jose Mena, sued Darin L. Muehler, the City of Simi Valley, California, and several law enforcement officers under 42 U.S.C. § 1983, alleging civil rights violations arising from a search of their home.[1] During the search, police detained the plaintiff for two to three hours with her hands cuffed behind her back[2], and an immigration officer questioned her about her immigration status.[3] The plaintiff alleged that her Fourth Amendment

---

1. Muehler v. Mena, 544 U.S. 93, 95-96 (2005). At 7 a.m. on February 3, 1998, Simi Valley Police executed a valid warrant to search the Menas' house and surrounding premises for, "among other things, deadly weapons and evidence of gang membership." *Id.* Jose Mena rented space on the property to several unrelated individuals, and police had reason to believe that at least one of the tenants was a member of the West Side Locos, a gang comprised mostly of illegal immigrants. *Id.* at 95-96; *id.* at 106 (Stevens, J., concurring). The targeted tenant, Raymond Romero, was the primary suspect in a recent gang-related drive-by shooting and was considered to be armed and dangerous. *Id.* at 106 (Stevens, J., concurring). At the same time the Menas' house was being searched, police executed a search warrant at Romero's mother's house and found him there. *Id.* at 107. Romero was cited for possession of marijuana and was released. *Id.*

2. *Id.* at 103 (Kennedy, J., concurring). Iris Mena awoke as the Simi Valley SWAT team entered her bedroom and handcuffed her at gunpoint. *Id.* at 96 (majority opinion). Police led the barefooted Mena out to a cold garage where she was detained for two to three hours with her hands cuffed while her house was searched. *Id.* at 107 (Stevens, J., concurring); Mena v. Simi Valley, 332 F.3d 1255, 1260 (9th Cir. 2003). Throughout her detention, Mena was not told why she was being detained, and she was "unarmed, docile, and cooperative in every respect." *Mena*, 332 F.3d at 1260, 1263. Two officers guarded Mena and three other handcuffed detainees in the garage; however, one of the guards was sent home after offering to join in the search rather than guard the detainees. *Muehler*, 544 U.S. at 103 (Kennedy, J., concurring); *id.* at 110 (Stevens, J., concurring). The search of the house yielded a pistol, some gang paraphernalia, and some marijuana; however, no contraband was found on Mena's person or in her bedroom. *Id.* at 96 (majority opinion); *id.* at 110 (Stevens, J., concurring).

3. *Id.* at 96 (majority opinion). After the immigration officer was informed that documents confirming Mena's immigration status were in her purse, a police officer proceeded to search her purse without consent. *Mena*, 332 F.3d at 1260.

rights were violated because her detention was conducted "for an unreasonable time and in an unreasonable manner."[4]

In the United States District Court for the Central District of California, a jury found that the plaintiff's Fourth Amendment rights were violated when she was detained "with unreasonably excessive force and for a longer period than was reasonable."[5] On appeal, the Ninth Circuit affirmed the district court's ruling that Mena was subjected to an unreasonable seizure.[6] On certiorari to the United States Supreme Court, *held*, vacated and remanded.[7] A law enforcement officer's power to detain an individual during a proper search is "categorical,"[8] and questioning a detainee about his citizenship status does not violate that individual's Fourth Amendment rights because such "questioning does not constitute a seizure."[9]

Section 1983 imposes civil liability for violating another's constitutional right to be free from unreasonable searches and seizures.[10] In assessing the scope of Fourth Amendment rights, courts have attempted to strike a balance between the protection of individual citizens from intrusions and the preservation of law enforcement interests. Over the last forty years, the trend has been to afford law enforcement officers an increasing amount of leeway in conducting searches and seizures.[11] In *Muehler v. Mena*,[12] the U.S. Supreme Court continued this trend of galvanizing police officers' authority during detentions by addressing two key issues: whether, in the course of executing a search warrant, a police officer has the authority to detain an occupant of the premises in handcuffs for the duration of the search, and whether police may

---

4. *Muehler*, 544 U.S. at 96.

5. *Mena*, 332 F.3d at 1260.

6. *Id.* at 1259. The Court of Appeals affirmed the district court on two grounds. First, the court held that Mena's detention in handcuffs was "objectively unreasonable," and that she should have been released when it became apparent that she posed no immediate threat. *Id.* at 1263. Second, the Court of Appeals held that the immigration officer's inquiry into Mena's immigration status constituted an independent violation of her constitutional right to privacy. *Id.* at 1266.

7. 544 U.S. at 102.

8. *Id.* 98.

9. *Id.* at 101 (quoting Florida v. Bostick, 501 U.S. 429, 434 (1991).

10. 42 U.S.C. § 1983 (2000). The Statute provides in pertinent part:

> Every person who, under color of any statute, ordinance, regulation, custom, or usage, of any State . . . subjects, or causes to be subjected, any citizen of the United States or other person within the jurisdiction thereof to the deprivation of any rights, privileges, or immunities secured by the Constitution and laws, shall be liable to the party injured in an action at law [or a] suit in equity . . . ."

Id.

11. *See, e.g.*, Michigan v. Summers, 452 U.S. 692, 704-05 (1981); Terry v. Ohio, 392 U.S. 1, 30-31 (1968).

12. 544 U.S. 93 (2005).

question a detainee about his citizenship status when they have no reason to suspect that the detainee is an illegal alien.[13]

II. PROBABLE CAUSE AND THE DEVELOPMENT OF THE OBJECTIVE REASONABLENESS STANDARD

On the eve of the American Revolution, American colonists began to rebel against British authority by adopting safeguards against unreasonable searches and seizures.[14] After the Revolution, the Fourth Amendment solidified the right of citizens to be free from "unreasonable searches and seizures" and established that "probable cause" must exist before such intrusions occur.[15] Similarly, contemporaneous judicial decisions disposed of the notion that mere suspicion of criminal activity was sufficient justification for a search or arrest.[16] These basic guarantees, which limit the circumstances in which the government may search or seize a citizen, have been upheld through the modern era.[17]

In *Terry v. Ohio*,[18] however, the Supreme Court departed from the general rule that no unreasonable search or seizure shall take place without probable cause.[19] The Court in *Terry* adopted the lower standard of "reasonable suspicion" for searches that constitute less of an intrusion than an arrest.[20] This change in the law opened the door for additional intrusions that could potentially occur incident to a search based on probable cause.

In the latter part of the 20th century, the Supreme Court was asked to address the reasonableness standard implicit in the Fourth Amendment.[21] In *Tennessee v. Garner*,[22] the Supreme Court held that "reasonableness depends on not only when a [search or] seizure is made, but also how it is carried out."[23] The Court's decision in *Garner*, however, was not the first time that the Court

---

13. *Id.* at 97.

14. *See* Henry v. United States, 361 U.S. 98, 100 (1959) (citing THE VIRGINIA DECLARATION OF RIGHTS (U.S. 1776)). In *Henry*, the court noted that the Declaration forbade the "'grievous and oppressive'" practice of granting search warrants when the suspected offense was "not particularly described and supported by evidence." *Id.*

15. U.S. CONST. amend. IV.

16. *See, e.g.*, Frisbie v. Butler, 1 Kirby 213, 215 (Conn. 1787).

17. *See e.g., Henry*, 361 U.S. at 104 (holding that a criminal conviction that is based on the unreasonable seizure of evidence must be overturned when there is no probable cause at the time the seizure is made).

18. 392 U.S. 1 (1968).

19. *Id.* at 30-31. Although the standard of probable cause for justifying an arrest has remained firm, the Court adopted the lower standard of "reasonable suspicion" for searches that constitute less of an intrusion than an arrest. *Id.* at 27, 30-31.

20. *Id.* at 27, 30-31.

21. *See, e.g.*, Tennessee v. Garner, 471 U.S. 1, 7-8 (1985); Michigan v. Summers, 452 U.S. 692, 694 (1981).

22. 471 U.S. 1 (1985).

23. *Id.* at 8.

had considered the reasonableness of a search or seizure. In *Michigan v. Summers*,[24] the Court first dealt with the propriety of police actions during the execution of a lawful search warrant.[25] The Court held that it was reasonable and appropriate for police "to detain the occupants of the premises while a proper search is conducted."[26] The primary rationale underlying this holding was to protect police officers and better enable them to prevent suspects from fleeing or resisting arrest.[27] The Court recognized that "the 'physical entry of the home is the chief evil against which the wording of the Fourth Amendment is directed.'"[28] However, the Court downplayed the detention of the plaintiff by noting that his "detention . . . was surely less intrusive than the [property] search itself."[29] The Court added that detaining occupants during a search provides those detainees with an incentive to aid in the search to avoid the use of force against their person or property.[30]

Although *Summers* extended the powers of law enforcement officers during the execution of a search warrant, the decision articulated a warning against unreasonable use of that power.[31] *Summers* limited the authority to detain occupants during the search of a premises by holding that law enforcement officers conducting a lawful search could still violate the Fourth Amendment under "special circumstances, or [in the event of] a prolonged detention."[32]

The delicate nature of the holding in *Summers* and the rapidly evolving law of search and seizure prompted the Court to revisit the balancing test used to

---

24. 452 U.S. 692 (1981).

25. *Id.* at 694-95. The plaintiff was exiting the house that police had a warrant to search when he was detained. *Id.* at 693. The plaintiff was arrested after police discovered narcotics in his house. *Id.* Police officers also discovered more narcotics in the plaintiff's coat pocket after "search[ing] his person." *Id.*

26. *Id.* at 705. This rule has been upheld and extended to include the removal of passengers from automobiles during searches. *See* Maryland v. Wilson, 519 U.S. 408, 415 (1997) (holding that during a traffic stop, a police officer may order passengers out of the vehicle "pending [the] completion of the stop"). The *Wilson* court based its holding on the premise that the additional intrusion upon passengers would be minimal because the only difference for passengers would be that they are outside of the car instead of inside the car. *Id.* at 414.

27. *Summers*, 452 U.S. at 702.

28. *Id.* at 701 n.13 (quoting United States v. U. S. Dist. Court for the E. Dist. of Mich., 407 U.S. 297, 313 (1972)).

29. *Id.* at 701. Along with this reasoning, the Court offered two other justifications for detaining suspects during a search of their property. *Id.* at 702. The Court assumed that, "unless they intended flight to avoid arrest," most citizens would want to remain inside their residence to watch the search of their belongings. *Id.* at 701. Even though the plaintiff exhibited no signs of violence or hostility, the Court added that the police should "exercise unquestioned command of the situation" so as to prevent any "sudden violence or frantic efforts to conceal or destroy evidence." *Id.* at 702-03.

30. *Id.* at 703.

31. *Id.* at 703-04.

32. *Id.* at 705 n.21.

determine the propriety of additional intrusions upon privacy.[33] In determining when a seizure based on less than probable cause should be allowed, the Court noted that it must balance the nature of the intrusion upon the individual's Fourth Amendment rights against the "governmental interests alleged to justify the intrusion."[34] Thus, the reasonableness of the manner in which a search or seizure is carried out "depends on a balance between the public interest and the individual's right to personal security free from arbitrary interference by law officers."[35]

In *Graham v. Connor*,[36] the Court examined the question of when excessive use of force incident to a proper seizure gives rise to a cause of action under § 1983.[37] The Court determined that the propriety of force used by police in a search or seizure, and therefore, the potential for § 1983 liability, should turn on whether the officer's actions were "'objectively reasonable.'"[38] This standard does not consider the officers' "underlying intent or motivation."[39] The Court also emphasized that the "[objective] 'reasonableness' of a particular use of force must be judged from the perspective of a reasonable officer on the scene," and not in hindsight.[40] Thus, the correct application of the test requires that the circumstances that may affect a finding of reasonableness with regard to the officer's actions be considered in their totality.[41]

Recently, the Ninth Circuit has been accused of mounting a "classic frontal assault" on the search and seizure precedents set forth by the Supreme Court.[42] In *Franklin v. Foxworth*,[43] the Ninth Circuit questioned the manner in which a search warrant was executed.[44] The court noted that determining the propriety

---

33. United States v. Place, 462 U.S. 696, 703 (1983).

34. *Id.*

35. United States v. Brignoni-Ponce, 422 U.S. 873, 878 (1975).

36. 490 U.S. 386 (1989).

37. The plaintiff, a diabetic, was seen frantically exiting a convenience store while trying to purchase orange juice to drink in order to remedy an insulin reaction. *Id.* at 388-89. A police officer who observed the plaintiff "hastily enter and leave the store" became suspicious and forcibly detained and handcuffed him. *Id.* at 389.

38. *Graham*, 490 U.S. at 397 (quoting Scott v. United States, 436 U.S. 128, 138 (1978)).

39. *Id.* The Court notes the reflexive effect of this standard where "[a]n officer's evil intentions will not make a Fourth Amendment violation out of an objectively reasonable use of force; nor will an officer's good intentions make an objectively unreasonable use of force constitutional." *Id.*

40. *Id.* at 396.

41. *Id.* For the objective reasonableness test to be correctly applied, the court must consider "the severity of the crime at issue, whether the suspect poses an immediate threat to the safety of the officers or others, and whether he is actively resisting arrest or attempting to evade arrest by flight." *Id.*

42. Ernest Harper, *Defending the Citadel of Reasonableness: Search and Seizure in 2004*, 2005 ARMY LAW 47, 49 (2005).

43. 31 F.3d 873 (9th Cir. 1994).

44. *Id.* at 875. The police executed a valid warrant to search the plaintiff's dwelling. *Id.*

of detention during a lawful search becomes more complex when the person being detained is not suspected of a crime.[45] The *Franklin* court also reasserted the *Graham* factors for finding an objectively unreasonable search, adding that unreasonableness may also be found where a search is "unnecessarily painful, degrading, or prolonged, or if it involves an undue invasion of privacy."[46] In *Franklin*, the court held that the police violated the warning set forth in *Summers*.[47] Although the plaintiff was detained legally for the duration of the search, the manner in which he was detained violated the Fourth Amendment.[48]

The Supreme Court's decisions in *Terry* and *Summers* support the position that law enforcement officers may detain a suspect while executing a search warrant.[49] However, the decisions in *Place*, *Garner*, and *Graham* supply a mandate that these searches and seizures be objectively reasonable, as scrutinized through a highly fact-intensive inquiry.[50]

Since its decision in *Terry*, the Supreme Court has had to determine what kind of interaction between citizens and police constitutes a search and seizure requiring Fourth Amendment scrutiny.[51] In *Terry*, the Court held that "whenever a police officer accosts an individual and restrains his freedom to walk away, he has 'seized' that person."[52] The Court pointed out "that the Fourth Amendment governs 'seizures'. . . which do not eventuate in a trip to the station house . . . ."[53] If a seizure does take place, a Fourth Amendment violation will result if the police officer is unable "to point to specific and articulable facts which, taken together with rational inferences from those facts, reasonably warrant that intrusion."[54] In a later case, the Supreme Court

---

at 874. The plaintiff, Curry, suffered from an advanced stage of Multiple Sclerosis and was "unable to walk." *Id.* at 874. Curry was handcuffed at gunpoint and carried half-naked to a central room where he remained for two hours while the house was searched. *Id.* at 875. The court held that a § 1983 violation had occurred because the police "subjected an obviously ill and incapacitated person to entirely unnecessary and unjustifiable degradation and suffering." *Id.* at 878.

45. *See id.* at 877. Parties who are present at a place being searched may be pursuing an "innocuous if not benevolent purpose" but can still be detained by appropriate means. *Id.* at 876.

46. *Id.*

47. *Id.*

48. *Id.* at 878. *Foxworth* was deemed to be "an example of the unusual case contemplated by *Summers*." *Id.* at 876.

49. *See* Michigan v.Summers, 452 U.S. 692, 703 (1981); Terry v. Ohio, 392 U.S. 1, 29 (1967).

50. United States v. Place, 462 U.S. 696, 703-10 (1983); Tennessee v. Garner, 471 U.S. 1, 7 (1984); Graham v. Connor, 490 U.S. 386, 395 (1989).

51. *See supra* notes 48-49.

52. *Terry*, 392 U.S. at 16.

53. *Id.*

54. *Id.* at 21. While this standard does not require that the officer have probable cause to stop an individual, an "articulable" reasonable suspicion is required. *Id. See also* Brown v. Texas, 443 U.S. 47, 52 (1979) ("In the absence of any basis for suspecting [someone] of

explained that its decision in *Terry* was designed to enable a police officer to determine the identity of a suspect "or to maintain the status quo momentarily" when that officer "lacks the precise level of information necessary for probable cause to arrest . . . ."[55]

In *United States v. Brignoni-Ponce*,[56] the Court articulated additional factors to be considered when determining whether the reasonable suspicion requirement for a *Terry* stop has been satisfied.[57] The Court noted that a seizure to inquire about a suspect's citizenship status is inappropriate when based solely on the fact that the suspect appears to be of Mexican ancestry.[58] Additionally, the Court pointed out that the protection against unreasonable seizures also extends to "seizures that involve only a brief detention short of traditional arrest."[59]

In *INS v. Delgado*,[60] the Court applied the "objectively reasonable" rule from *Terry* to an encounter in which the alleged seizure consisted only of questioning by law enforcement officers.[61] In *Delgado*, the Court looked objectively at the impressions of the person being questioned to determine whether a Fourth Amendment violation occurred.[62] The Court echoed the *Terry* definition of a seizure by holding that "[u]nless the circumstances of the encounter are so intimidating as to demonstrate that a reasonable person would have believed he was not free to leave if he had not responded," then there is no "detention under the Fourth Amendment."[63] Under *Delgado*, mere

---

misconduct, the balance between the public interest and [the person's] right to personal security and privacy tilts in favor of freedom from police interference.").

55. Adams v. Williams, 407 U.S. 143, 145, 146 (1972). *See also* Delaware v. Prouse, 440 U.S. 648, 653 (1979) (holding that a seizure can occur even when "the purpose of the stop is limited and the resulting detention [is] quite brief").

56. 422 U.S. 873 (1975).

57. *Id.* at 884-87. At issue in this case was the propriety of a stop made by border patrol agents in an area near the Mexican border. *Id.* at 874. While scanning the highway for illegal immigrants, the agents stopped the car in which the plaintiffs were passengers solely because the plaintiffs "appeared to be of Mexican descent." *Id.* at 875. The agents then questioned the plaintiffs about their citizenship and determined that they were, in fact, illegal aliens. *Id.* The Court reemphasized that law enforcement officers may not stop a vehicle on "a random basis," and that officers must have a "reasonable suspicion for stops." *Id.* at 883.

58. *Id.* at 886-87.

59. *Id.*at 878

60. 466 U.S. 210 (1984).

61. *Id.* at 215. INS agents entered a garment factory pursuant to search warrants issued on the probable cause that several illegal immigrants worked in the factory. *Id.* at 212. Several agents systematically questioned each worker about his or her immigration status, while several other agents positioned themselves by the factory's exits. *Id.* The plaintiffs alleged that the agents' actions "violated their [constitutional] right to be free from unreasonable [search and seizure]." *Id.* at 213.

62. *Id.* at 220-21.

63. *Id.* at 216.

questioning by law enforcement officers does not constitute a Fourth Amendment seizure.[64]

In *Florida v. Bostick*,[65] the Court evaluated additional circumstances surrounding police questioning that could draw Constitutional scrutiny.[66] In *Bostick*, the police boarded the bus in which Bostick was riding and, without reasonable suspicion, questioned him and requested consent to search his baggage.[67] Bostick consented to the search, and the police found cocaine in his baggage.[68] The Court noted that an "encounter will not trigger Fourth Amendment scrutiny unless it loses its consensual nature."[69] The "consensual nature" of an encounter may be lost and a seizure may occur if an officer, through "physical force or show of authority, has in some way restrained the liberty of a citizen."[70] Bostick argued that his encounter with the police rose to the level of a seizure because it happened in the "cramped confines of a bus," where the police were very intimidating and gave him the impression that he was not free to leave.[71] The Court held that there should be no *per se* rule that dictates when an encounter is so intimidating that it becomes a seizure.[72] However, the Court added that facts such as those asserted by Bostick are relevant in the fact-intensive inquiry of determining whether a Fourth Amendment seizure has occurred.[73]

---

64. *Id.* at 221. *See also* Florida. v. Royer, 460 U.S. 491, 497-98 (1983) (stating that a reasonable person may turn his or her back on a police officer's questions and, thereby, demonstrate the freedom that prevents such an encounter from falling under Fourth Amendment scrutiny).

65. 501 U.S. 429 (1991).

66. *Id.* at 433. The Court stressed that "all of the circumstances surrounding [an] encounter" must be considered when determining whether the interaction between and individual and the police is a consensual encounter or a Fourth Amendment seizure. *Id.* at 437.

67. *Id.* at 431-32. The Court took special note of the fact that "the police specifically advised Bostick that he had the right to refuse consent." *Id.* at 432.

68. *Id.* at 431.

69. *Id.* at 434.

70. *Id.* (quoting Terry v. Ohio, 392 U.S. 1, 19 n.16 (1968)). This holding, which follows the holding in *Delgado*, illustrates the type of intimidation that the Court foresees as causing a reasonable person to feel that an encounter is non-consensual.

71. *Id.* at 435.

72. *See id.* at 440.

73. *Id.* at 439. Furthermore, the Court made the important distinction that for a seizure to be based on a suspect's objectively reasonable belief that he is not free to leave, it must be the force of law enforcement officials that is restricting the suspect's movement. *Id.* at 437. The Court noted that Bostick may not have felt free to leave even if police were not present because he had already lodged himself in the cramped confines of the bus. *Id.* at 436. The Court added that this point of law fits within the holding of *Delgado* because, even though the agents in that case stood at the exits of the factory, it was the worker's employment responsibilities and not the agents' actions that kept the workers from leaving the building. *Id.*

The Court recently expanded the search and seizure powers of law enforcement with its decision in *Illinois. v. Caballes*.[74] The Illinois Supreme Court held in *Caballes* that a Fourth Amendment violation occurred because there were no reasonable "articulable facts"justifying a search for drugs while the defendant was detained for a speeding violation.[75] However, the United States Supreme Court held that the use of drug-sniffing dogs did not constitute a search and was lawful because it did not prolong the "time reasonably required" to complete the already proper stop for speeding.[76] The Court also noted that the dog-sniff of Caballes' trunk did not constitute a search and was not an unreasonable invasion of privacy because the dog-sniff would only reveal illegal behavior and would not divulge any lawful activity that Caballes would wish to keep private.[77] Thus, if there is a seizure based on reasonable suspicion that gives rise to a further investigation that is not based on articulable facts, then that additional investigation may still be lawful as long as it does not prolong the original seizure.

The Supreme Court's decisions in *Terry*, *Brignoni-Ponce*, *Delgado*, and *Bostick* establish the elements to be used in determining whether an encounter involving police questioning requires Fourth Amendment scrutiny. The general guidance offered by the court is that police questioning alone does not constitute a seizure.[78] However, it is significant that the Court has opened the door for a case where police questioning could constitute a seizure, depending upon the circumstances surrounding the police interaction with the suspect.[79]

III. *MUEHLER V. MENA*: CONTINUED EXPANSION OF POLICE SEARCH AND SEIZURE POWERS

In *Muehler v. Mena*,[80] the Supreme Court confirmed that law enforcement officers have broad authority to detain individuals incident to a search and that police do not need independent reasonable suspicion to question detainees

74. 543 U.S. 405, 410 (2005). The defendant, Caballes, was stopped for a speeding violation. *Id.* at 406. While the officer that pulled Caballes over was writing him a ticket, another officer arrived on the scene and walked his drug-sniffing dog around Caballes' car. *Id.* The dog detected the presence of marijuana, which was subsequently found in Caballes' trunk. *Id.* The court noted that this "entire incident lasted less than 10 minutes." *Id.*

75. *Id.* at 407. The state supreme court held that "the use of the dog 'unjustifiably enlarg[ed] the scope of a routine traffic stop into a drug investigation.'" *Id.* (quoting Illinois v. Caballes, 802 N.E.2d 202, 205 (Ill. 2003)).

76. *Id.* at 407-409.

77. *Id.* at 409. The court drew a contrast between this case and Kyllo v. United States, 533 U.S. 27 (2001), where the court held that "the use of a thermal-imaging device to detect the growth of marijuana in a home" was unlawful because "the device was capable of detecting lawful activity . . . such as 'at what hour each night the lady of the house takes her daily sauna and bath.'" *Id.* at 409-10 (quoting Kyllo v. United States, 533 U.S. 27, 38 (2001)).

78. *Bostick*, 501 U.S. at 434.

79. *See, e.g., id.* at 439; INS v. Delgado, 466 U.S. 210, 220-21 (1984).

80. 544 U.S. 93 (2005).

about their citizenship status.[81] In light of these holdings, the Court ruled that no Fourth Amendment violation had taken place and, therefore, the police were not liable to Mena under § 1983.[82] The majority opinion represents a continuation of the Supreme Court's effort to expand the power of law enforcement by authorizing their seemingly intrusive search and seizure tactics.

In the first part of the majority opinion, Chief Justice Rehnquist overruled the Ninth Circuit's decision that Mena was detained for an unreasonable time and in an unreasonable manner.[83] Chief Justice Rehnquist confronted the Ninth Circuit's holding that Mena's detention was unreasonable because she was not released from her handcuffs "as soon as it became clear that she posed no immediate threat."[84] Although a jury found that Mena's rights had been violated, Chief Justice Rehnquist made it clear that the Court does "not defer to the jury's legal conclusion" about a violation of the Constitution.[85]

Chief Justice Rehnquist closely followed the Supreme Court's decision in *Summers*, holding that "[a]n officer's authority to detain incident to a search is categorical."[86] Chief Justice Rehnquist operated on the precept from *Summers* that the detention of an occupant of a searched house is "'surely less intrusive than the search itself.'"[87] Thus, the majority in *Muehler* held that the intrusion upon Mena was "slight" and justified by the interests of law enforcement.[88]

Chief Justice Rehnquist echoed several of the law enforcement interests expressed in *Summers* that justified Mena's detention.[89] The intrusion on Mena was outweighed by the governmental interest in "'preventing flight in the event that incriminating evidence is found'; 'minimizing the risk of harm to the officers'; and facilitating 'the orderly completion of the search,' as detainees' 'self interest may induce them to open locked doors . . . to avoid the use of force.'"[90] The Court also found that the use of handcuffs in Mena's case was reasonable because police interests "are at their maximum when . . . a warrant authorizes a search for weapons and a wanted gang member resides on the premises [to be searched]."[91] The Court also considered the fact that the detainees outnumbered the officers who were guarding them to be an important

---

81. *Id.* at 100-01.

82. *Id.* at 102.

83. *Id.* at 95.

84. *Id.* at 97.

85. *Id.* at 98 n.1.

86. *Id.* at 98. According to *Summers*, "[a]n officer's authority to detain . . . does not depend on 'the quantum of proof justifying detention or the extent of the intrusion to be imposed by the seizure.'" *Id.* (quoting Michigan v. Summers, 452 U.S. 692, 705 n.19 (1981).

87. *Id.* (quoting *Summers*, 452 U.S. at 701).

88. *Id.*

89. *Id.* (quoting *Summers*, 452 U.S. at 702-03).

90. *Id.*

91. *Id.* at 100.

justification for handcuffing Mena.[92] In light of the officers' "categorical" authority to detain and the guidance offered by *Summers*, the Court held that Mena's two to three hour detention in handcuffs was "plainly permissible."[93]

The majority in *Muehler* also addressed whether the defendants violated the plaintiff's Fourth Amendment rights by questioning her about her immigration status during the detention.[94] In accordance with prior Supreme Court decisions, the Court affirmed that there was no constitutional violation because "'mere police questioning does not constitute a seizure.'"[95] The Court explained that this holding allows police to "'generally ask questions of [an] individual; ask to examine the individual's identification; and request consent to search his or her luggage.'"[96] The Court added, however, that police questioning may transform into a seizure if the questioning is unduly prolonged.[97] In response to the Ninth Circuit's decision below, the Supreme Court emphatically held that there is no "'requirement of particularized reasonable suspicion for . . . inquiry into citizenship status.'"[98]

Writing separately, Justices Kennedy and Stevens each concurred in the majority opinion.[99] They expressed concern, however, that the majority opinion could lower the bar of objective reasonableness to the point where police would be emboldened to use search and seizure tactics that could slip into the realm of unconstitutionality.[100]

Justice Kennedy agreed that Mena's detention was properly conducted and, therefore, reasonable.[101] Justice Kennedy's opinion was based on deference for the issues of officer safety and efficiency of the search.[102] Like the majority, Justice Kennedy found the detention reasonable because the "detainees outnumber[ed] those supervising them."[103]

Although Justice Kennedy agreed with the majority decision, he would have preferred that Chief Justice Rehnquist include some directive dicta to better ensure the reasonableness of police actions in future detentions.[104] Justice Kennedy's chief concern was to "ensure that police handcuffing during

---

92. *Id.*
93. *Id.* at 98.
94. *Id.* at 100-01.
95. *Id.* at 101 (quoting Florida v. Bostick, 501 U.S. 429, 434 (1991); *see also* INS v. Delgado, 466 U.S. 210, 212 (holding that a factory survey by INS does not constitute a seizure under the Fourth Amendment).
96. *Muehler*, 544 U.S. at 101 (quoting *Bostick*, 501 U.S. at 435).
97. *Id.*
98. *Id.* at 101-02 n.3 (quoting Mena v. Semi Valley, 332 F.3d 1255, 1267) (9th Cir. 2003).
99. *Id.* at 102, 104 (Kennedy, J., concurring).
100. *Id.* at 102-06.
101. *Id.* at 103.
102. *Id.*
103. *Id.*
104. *Id.* at 102.

searches becomes neither routine nor unduly prolonged."[105] Contrary to the majority's emphasis on the "categorical"[106] authority of police to detain, Justice Kennedy would remind law enforcement officers that they are constantly subject to the "reasonableness calculation under *Graham*."[107] Justice Kennedy added that a detainee's handcuffs should be removed if "it would be readily apparent to [an] objectively reasonable officer" that doing so "would not compromise the officers' safety or risk interference or substantial delay in the execution of the search."[108] Furthermore, even if continued handcuffing is necessary, Justice Kennedy asserted that a detainee could be temporarily released "under supervision to avoid pain or excessive physical discomfort."[109]

In a separate opinion, Justice Stevens concurred with the majority's holding, but also recognized the slippery slope that could arise from Chief Justice Rehnquist's opinion.[110] Justice Stevens agreed that the case should be remanded to determine whether the police questioning about Mena's citizenship status caused her to be detained longer than the authorized search lasted.[111] However, Justice Stevens disagreed with the majority's interpretation of *Summers* and *Graham*.[112] Justice Stevens argued that *Summers* authorizes only the detention of someone who is present when a warrant is executed; it "does not give officers *carte blanche* to keep individuals who pose no threat in handcuffs throughout a search."[113] Justice Stevens advocated a narrower reading of *Summers*—one that does not give police "categorical" discretion over the force used in a detention.[114] Justice Stevens pointed out that the plaintiff could have been handcuffed in many ways that would have reduced her discomfort but maintained security.[115] Nevertheless, Justice Stevens agreed with the majority that all occupants of a house may be handcuffed when a search is initially executed because police should not have to assess who is a threat prior to handcuffing.[116] However, Justice Stevens expressed an expectation that, once the dust settles, the "objective reasonableness test" should prevail.[117] Justice Stevens also asserted that the majority misinterpreted the objective reasonableness test with respect to the charge that the Court gave to the Ninth Circuit on remand.[118] Justice Stevens stressed that the issue to be

105. *Id.*
106. *Id.* at 98 (majority opinion).
107. *Id.* at 103 (Kennedy, J., concurring).
108. *Id.*
109. *Id.* at 103-04.
110. *Id.* at 105-06 (Stevens, J., concurring).
111. *Id.* at 106.
112. *Id.* at 105.
113. *Id.* at 105-06 (emphasis in original).
114. *Id.*
115. *Id.* at 109 n.6.
116. *Id.*
117. *Id.* at 110.
118. *Id.* at 111-12.

evaluated on remand regarding objective reasonableness is "whether the evidence supports Mena's contention that the petitioners used excessive force in detaining her . . . ."[119]

The Supreme Court's decision in *Muehler v. Mena* represents a regrettable erosion of the protections of the Constitution. In an effort to advance the powers of law enforcement, the Court misinterpreted legal precedent and overlooked key facts while downplaying the importance of the objective reasonableness standard. The Court incorrectly analyzed the issue of the manner of Mena's detention. The "categorical" authority that the Court gives police officers in this decision threatens to swallow the objective reasonableness analysis that should govern police actions.[120] Furthermore, the majority failed to consider many mitigating facts that were pertinent to the issue of whether Mena should have been kept in handcuffs.

On the issue of police questioning, the court correctly held that "'mere police questioning does not constitute a seizure.'"[121] However, the court failed to consider the circumstances surrounding Mena's questioning that could have made the detention qualify as a seizure under the Fourth Amendment. The Court misinterpreted the authority that *Summers* gives to police in executing a search warrant founded on probable cause. *Summers* clearly authorizes police to detain the occupants of the premises being searched for the duration of the search.[122] However, the Court in *Muehler* interpreted *Summers* to allow police to "'exercise unquestioned command of the situation'" during a search, as opposed to granting police broad discretion in deciding which occupants to detain throughout the search.[123] The Supreme Court noted the language from *Summers* stating that the risk of injury to all involved is "minimized 'if the officers routinely exercise unquestioned command of the situation.'"[124] Although the safety of police officers during a search is of great importance, the Court's holding is potentially dangerous because police actions should always face the scrutiny of an impartial magistrate.[125] By using language that suggests that the police will not be questioned in the use of their "categorical"[126] authority, the Court gave the police license to make "ad hoc legal judgments which pose a serious threat to Fourth . . . Amendment protections."[127]

The majority overlooked several circumstances surrounding Mena's detention that could have supported a verdict that the continued use of

---

119. *Id.* at 106.

120. *Id.* at 98 (majority opinion).

121. *Id.* at 101 (quoting Florida v. Bostick, 501 U.S. 429, 434 (1991)).

122. Michigan v. Summers, 452 U.S. 692, 705 (1981).

123. *Muehler*, 544 U.S. at 99 (quoting *Summers*, 452 U.S. at 703).

124. *Id.*

125. *See* Terry v. Ohio, 392 U.S. 1, 21 (1967) (noting that "the Fourth Amendment becomes meaningful only when it is assured that at some point the conduct of those charged with enforcing the laws can be subjected to the more detached, neutral scrutiny of a judge").

126. *Muehler*, 544 U.S. at 98.

127. *Summers*, 452 U.S. at 712 n.5 (Stewart, J., dissenting).

handcuffs was objectively unreasonable. The Court reached its holding that Mena's handcuffing was reasonable by focusing on the fact that the detainees outnumbered the officers guarding them.[128] In reaching this holding, the majority ignored the fact that one of the police officers guarding Mena "was sent home after offering to assist in the search."[129] This fact suggests that an objectively reasonable officer on the scene felt that Mena was such a minor threat that it would be reasonable to release one of the officers guarding her.[130] The Court also found that the police acted reasonably because the warrant called for a search for weapons at a location where a "wanted gang member reside[d]."[131] However, this rationale ignores the fact that Romero, the alleged gang member, was not wanted as a dangerous criminal; otherwise, it is unlikely that the police would have released him that same day after citing him for marijuana possession only.[132]

Furthermore, given Mena's circumstances, the Court operated on a flawed justification for keeping her in handcuffs.[133] Mena was not in a position to flee "in the event that incriminating evidence [was] found" because she was detained in a garage where she could not observe the search.[134] Additionally, it seems unreasonable to think that Mena, an unarmed woman standing five feet, two inches tall, could pose a serious threat to the safety of an armed police officer.[135] In Mena's case, her detention was not "surely less intrusive than the search itself."[136] Being held for hours in a cold garage with one's hands cuffed behind the back is surely just as unpleasant, if not more unpleasant, than having one's possessions inspected. In its limited analysis of the police questioning issue, the Court failed to follow its own precedent. The Court correctly held that simple police questioning does not constitute a seizure, but failed to consider other facts that could give rise to the finding of a seizure. The Court correctly noted that the questioning of Mena would constitute a seizure if it "prolonged . . . the time reasonably required" to search the premises.[137] However, the Court did not consider whether, in light of the circumstances, "the encounter [was] so intimidating as to demonstrate that a reasonable person would have believed he was not free to leave if he had not responded . . . ."[138]

---

128. *Muehler*, 544 U.S. at 100.

129. *Id.* at 110 (Stevens, J., concurring).

130. *See Terry*, 392 U.S. at 21-22 (holding that the reasonableness of police action "be judged against an objective standard: would the facts available to the officer at the moment of the seizure or the search" cause him to believe that the appropriate action was being taken?).

131. *Muehler*, 544 U.S. at 100.

132. *Id.* at 107 (Stevens, J., concurring).

133. *Id.* at 98-100 (majority opinion).

134. *Id.* at 98 (quoting Michigan v. Summers, 452 U.S. 692, 702 (1981)).

135. *Id.* at 105 (Stevens, J., concurring).

136. *Id.* at 98 (majority opinion) (quoting *Summers*, 452 U.S. at 701).

137. *Id.* at 101 (quoting Illinois v. Caballes, 543 U.S. 405, 407 (2005)).

138. INS v. Delgado, 466 U.S. 210, 216 (1984).

Considering the especially intrusive circumstances of Mena's detention, including Mena's being dragged out of bed at gunpoint, she surely could have felt intimidated by the police questioning and may have believed that she did not have the right to remain silent. Furthermore, unlike the detainees in *Bostick*, Mena's belief that she was not free to walk away from the questioning would have been entirely caused by the police activity.[139] Thus, the Court erred in overlooking these circumstances that could have caused Mena's questioning to qualify as a seizure.

Unfortunately, the decision in *Muehler* may have sent the law of search and seizure down a slippery slope. This decision may embolden police to exercise their "categorical" authority by using unreasonable tactics during searches and seizures.[140] After this decision, the reasonableness standard that exists to temper police searches may be lowered. Although granting law enforcement agents complete authority during searches and seizures may save lives, the Court must not lose sight of the reasonableness standard that pervades every governmental intrusion upon the rights of citizens. Otherwise, the Court will be promoting the evils that the Fourth Amendment was written to guard against. Additionally, there has been a trend in the post-September 11 world to broaden the powers of law enforcement in the name of security.[141] Accordingly, the fact that the Court found no impropriety in the officers' failure to tell Mena why she was being detained could make this decision very significant as issues arise concerning uncharged detainees in the War on Terrorism.

The *Muehler* decision represents the Supreme Court's desire to expand the power of law enforcement officers by lowering the Fourth Amendment's reasonableness standard. The Court went too far in declaring that the police have "categorical"[142] authority to detain during a search or seizure. The Court's declaration suggests that the police have unquestioned discretion in the tactics used to effect a detention. Although the Court correctly held that simple police questioning does not constitute a seizure, it failed to consider the circumstances surrounding Mena's questioning that could have qualified it as a seizure. Sadly, this decision advances the legal anomaly that, while a search warrant requires a finding of probable cause by a neutral magistrate, police officers may execute that warrant using tactics that will go virtually unquestioned.

BROCK BOSSON

---

139. *See* Florida v. Bostick, 501 U.S. 429, 436 (1991).

140. *Muehler*, 544 U.S. at 98.

141. *See, e.g.*, Michael P. O'Connor & Celia Rumann, *Going Going Gone: Sealing the Fate of the Fourth Amendment*, 26 FORDHAM INT'L L.J. 1234, 1249-55 (2003).

142. *Muehler*, 544 U.S. at 98.

PROPERTY—RIGHTS OF LAND OWNERS—OWNER'S RIGHT TO LATERAL SUPPORT OF LAND THAT HAS BEEN ALTERED FROM ITS NATURAL STATE

XI Props., Inc. v. RaceTrac Petroleum, Inc., 151 S.W.3d 443 (Tenn. 2004).

I. INTRODUCTION

In 1993, RaceTrac Petroleum, the defendant, sold 3.221 acres of undeveloped land to the plaintiff.[1] The deed stated that the land was sold "AS IS, WITH ALL FAULTS in its present condition, without guaranties or warranties, express or implied."[2] The land was adjacent to property that the defendant owned and operated as a filling station.[3] Prior to conveyance, the defendant added fill dirt to level an area near the filling station for a parking lot and to support the land.[4] In 1999, the plaintiff discovered that the defendant had inadvertently conveyed a portion of the land supporting the parking lot to the plaintiff.[5]

The plaintiff's plans for developing the property included removal of the slope supporting the defendant's parking lot.[6] The plaintiff sent the defendant a letter alerting the defendant of plans to remove the sloped embankment inadvertently sold to the plaintiff.[7] The defendant protested the plaintiff's plans.[8] The plaintiff filed a declaratory judgment action to clarify the parties' duties.[9] After both parties moved for summary judgment, the trial court granted the plaintiff's motion.[10] The court held that the land had been altered from its natural state when it was filled to construct the parking lot.[11] The plaintiff owed a duty of lateral support to land in its natural state; thus, the defendant was not entitled to support for the improvements.[12] The defendant appealed, claiming that the plaintiff purchased the tract "as is" and that the court should

---

1. XI Props., Inc. v. RaceTrac Petroleum, Inc., 151 S.W.3d 443, 445 (Tenn. 2004).
2. XI Props., Inc. v. RaceTrac Petroleum, Inc., No. M2001-00977-COA-R3-CV, 2003 Tenn. App. LEXIS 599, at *4 (Tenn. Ct. App. Aug. 27, 2003).
3. *XI Props., Inc.*, 151 S.W.3d at 445.
4. *Id.*
5. *Id.*
6. *Id.*
7. *Id.*
8. *Id.*
9. *Id.*
10. *Id.* at 446.
11. *Id.*
12. *Id.* at 449-50.

have ruled that the land reverted to its "natural state" when it was subdivided.[13] The appellate court affirmed the trial court ruling and, in addition, held that the plaintiff may not negligently excavate the property.[14] Upon review by the Tennessee Supreme Court, *held*, affirmed.[15] A landowner does not owe an absolute duty of lateral support to adjoining land that has been altered from its natural state, but a landowner who excavates land has a duty to exercise reasonable care to avoid unnecessary damage to adjoining property. *XI Props., Inc. v. RaceTrac Petroleum, Inc.*, 151 S.W.3d 443 (Tenn. 2004).

II. ISSUE

Landowners are entitled to expect that their land will not be put in danger by acts of their neighbors.[16] Tennessee recognizes that landowners owe a duty to provide lateral support to adjoining property in its natural state.[17] In *XI*, the Tennessee Supreme Court addressed whether landowners owe a duty of lateral support to adjoining land that has been altered from its natural state.[18] The court also resolved the duty of care that landowners owe adjoining property when excavating or improving their own land.[19]

III. DEVELOPMENT OF LATERAL SUPPORT LAW

Many states resolved the issue of whether there is a duty of lateral support to land that has been altered from its natural state over 100 years ago.[20] An

---

13. *Id.* at 447.

14. XI Props., Inc. v. RaceTrac Petroleum, Inc., No. M2001-00977-COA-R3-CV, 2003 Tenn. App. LEXIS 599, at *22 (Tenn. Ct. App. Aug. 27, 2003). The defendant asserted that it had been in possession of the land sold to the plaintiff for seven years and owned the property adversely. *Id.* at *5. The court held that the defendant did present issues of material fact, and the trial court was reversed on this issue and remanded to determine if all elements of adverse possession have been met. *Id.* at *15-16.

15. *XI Props., Inc.*, 151 S.W.3d at 445.

16. *See, e.g.*, 1 TENNESSEE JURISPRUDENCE *Adjoining Landowners* § 2 (2001).

17. *See* Williams v. S. Ry. Co., 396 S.W.2d 98, 99 (Tenn. Ct. App. 1965); Morris v. Ostertag, 376 S.W.2d 720, 723 (Tenn. Ct. App. 1963).

18. *XI Props., Inc.*, 151 S.W.3d at 446.

19. *Id.* at 448.

20. *See generally* Myer v. Hobbs, 57 Ala. 175 (1876) (applying English and American principles to develop the rights of lateral support); Green v. Berge, 38 P. 539 (Cal. 1894) (holding landowners strictly liable for excavations that caused damage to adjoining soil); Livingston v. Moingona Coal Co., 49 Iowa 369 (1878) (holding an excavator of coal liable for removing and failing to replace beams that were necessary to support the plaintiff's premises); Abrey v. City of Detroit, 86 N.W. 785 (Mich. 1901) (requiring a landowner who filled his property to erect a structure to keep the fill off neighboring land); Hutchinson & Rourke v. Schimmelfeder, 40 Pa. 396 (1861) (preventing landowners from injuring a neighbor's property); Stearns v. City of Richmond, 14 S.E. 847, 848 (Va. 1892) (noting the "ancient principle" that landowners have a right to lateral support from adjoining landowners).

issue of first impression in Tennessee, the court examined case law from other states in their holding. A brief examination of the development of lateral support laws in the United States provides a useful backdrop because the Tennessee Supreme Court relied on decisions from other states in reaching its decision in *XI*.[21]

A. Common Law Principles of Lateral Support in the United States

A fundamental principle of property law is that owners of land in its natural state have the right to lateral support from adjoining land.[22] Naturally necessary lateral support is "that support which the supported land itself requires and which, in its natural condition and in the natural condition of the surrounding land, it would require."[23] Lateral support is one of many sticks in the bundle of property rights.[24] A landowner's right to lateral support is fundamental to property because "in a state of nature all land is held together and supported by adjacent lands by operation of the forces of nature."[25] This country's "common law" of lateral support has developed over the course of two hundred years through well-supported cases from many states.[26] The principles of lateral support are so fundamental and logical that they have been used to develop other areas of the law.[27]

The first, and most elementary, principle of lateral support is that a landowner is entitled to naturally necessary lateral support from adjoining land.[28] In 1876, the Alabama Supreme Court in *Myer v. Hobbs*[29] held that landowners have the right to expect lateral support from adjoining land.[30] The court limited the right of lateral support to land that is in its natural state, and refused to hold landowners responsible for the support of adjoining property containing houses or other structures.[31] Since *Myer*, most courts have

---

21. *See XI Props., Inc.*, 151 S.W.3d at 447-50.

22. 1 AM. JUR. 2D *Adjoining Landowners* § 40 (2004) [hereinafter AM. JUR. 2D].

23. RESTATEMENT (SECOND) OF TORTS § 817 cmt. c (1979).

24. Myrl L. Duncan, *Reconceiving The Bundle of Sticks: Land as a Community-Based Resource*, 32 ENVTL. L. 773, 798-800 (2002).

25. Sime v. Jensen, 7 N.W.2d 325, 327 (Minn. 1942); *see also* Duncan, *supra* note 24, at 800 (stating that "the doctrine of lateral support emphasizes the inherently interconnected nature of private property").

26. *See* AM. JUR. 2D, *supra* note 22, § 40.

27. *See generally* Duncan, *supra* note 24, at 800 (discussing how lateral support provided a background for zoning and environment law); Christopher B. Amandes, Comment, *Controlling Land Surface Subsidence: A Proposal for a Market-Based Regulatory Scheme*, 31 UCLA L. REV. 1208, 1215-18 (1984) (discussing how lateral support theories in water subsidence actions aided in shaping modern subsidence laws).

28. *See, e.g.*, AM. JUR. 2D, *supra* note 22, § 40.

29. 57 Ala. 175 (1876).

30. *Id.* at 176.

31. *Id.* at 176-77; *see, e.g.*, Puckett v. Sullivan, 12 Cal. Rptr. 55, 58 (Cal. Ct. App. 1961); Massell Realty Improvement Co. v. MacMillan Co., 147 S.E. 38, 39 (Ga. 1929).

broadened this duty by holding landowners liable under certain circumstances for damages caused to structures on adjoining land.[32]

The second principle of lateral support is that landowners may excavate and improve their land as long as they do not cause injury to adjoining land in its natural state.[33] In *Carrig v. Andrews*,[34] the Connecticut Supreme Court held that landowners have the right to use their land in any way, as long as they provide artificial support to replace any natural support they remove.[35] A landowner even has the right to excavate "up to the very property line" as long as the support is replaced.[36]

The third principle of lateral support is that adjoining landowners are strictly liable for injuries caused by their removal of naturally necessary lateral support.[37] In *Levi v. Schwartz*,[38] the court held that if a landowner withdraws naturally necessary support from adjoining land, the damaged neighbor does not have to show negligence in order to recover damges.[39] The duty of lateral support is absolute.[40]

B. The Development of Lateral Support in Tennessee

Tennessee's lateral support doctrine was initially developed through cases involving party walls;[41] however, one case involving road construction paved the way for defining property owners' rights and duties with respect to neighboring land. In *Humes & Williams v. Mayor of Knoxville*,[42] the court found no legal remedy when an adjoining landowner's excavations injured the

---

32. *See, e.g.*, TENNESSEE JURISPRUDENCE, *supra* note 16, at § 3 ("If the landowner removes the soil from his own land so near the land of his neighbor that his neighbor's soil will crumble away under its own weight, he is liable for damages naturally resulting therefrom, including damages to structures on the subsiding land."); *cf.* RESTATEMENT, *supra* note 23, § 817 cmt. d (emphasizing that when land contains artificial additions, adjoining landowners owe a duty to provide "lateral support, and only that support, which the land itself requires, but not exceeding what it would require in the absence of these additions").

33. AM. JUR. 2D, *supra* note 22, § 47; *see also* Livingston v. Moingona Coal Co., 49 Iowa 369, 371 (1878) (stating that while the defendant landowner had a right to mine property for coal, the defendant was not permitted to remove so much coal that it deprived the plaintiff of the plaintiff's own property).

34. 17 A.2d 520 (Conn. 1941).

35. *Id.* at 521.

36. *Id.*; *accord* AM. JUR. 2D, *supra* note 22, § 47.

37. AM. JUR. 2D, *supra* note 22, § 48.

38. 95 A.2d 322 (Md. 1953).

39. *Id.* at 326.

40. *Id.*; *see, e.g.*, Green v. Berge, 38 P. 539, 541 (Cal. 1894); Law v. Phillips, 68 S.E.2d 452, 457 (W. Va. 1952).

41. *See, e.g.*, Carroll Blake Constr. Co. v. Boyle, 203 S.W. 945, 946 (Tenn. 1918); Murray v. Patterson, 72 S.W.2d 558, 562-64 (Tenn. Ct. App. 1934).

42. 20 Tenn. 1 Hum. 403 (1839).

plaintiff's property.[43] The property owners filed suit against city authorities alleging trespass on the property,[44] which was situated along a road owned by the city.[45] While improving the road for the public convenience, the city damaged the plaintiff's stable.[46] The lower court ruled against the plaintiff, finding no trespass.[47] The Tennessee Supreme Court upheld the lower court's ruling and held that the damage was not the result of intentional conduct or due to the city's failure to exercise reasonable care.[48] The court ruled that landowners have dominion over the air above and the soil below their property and may occupy it in any way, regardless of the consequences to adjoining landowners.[49] But landowners may not occupy the land in a manner that will cause intentional or foreseeable injuries to neighbors.[50] The *Humes & Williams* ruling remains in place for city officials regarding road construction. Tennessee courts would soon determine that such absolute permission to occupy land is not acceptable for all landowners.[51]

Tennessee recognized the common law duty of lateral support in *Carroll Blake Construction Co. v. Boyle*.[52] In that case, the plaintiffs and defendants shared a party wall that supported their respective buildings.[53] The defendants decided to tear down their building and construct a new building.[54] In order to provide support for the new building's foundation, the defendants had to excavate below the party wall.[55] The party wall collapsed, and the plaintiffs

---

43. *Id.* at 408.

44. *Id.* at 403. Plaintiffs alleged that the defendants

> [W]rongfully and unjustly, and without leave of and against the will of said plaintiffs, did cause the said Prince street, in front of and adjoining to said lot of ground, with the appurtenances thereon, as before described, to be dug down and moved away, so that thereby the foundations of said stable were greatly injured and impaired . . . .

Id. at 404.

45. *Id.*

46. *Id.* at 405.

47. *Id.* at 406.

48. *Id.* at 408.

49. *Id.* at 407. To illustrate this point, the court stated that if A owns a house on his property and adjoining landowner B decides to build a house, B may dig on his property even if it causes A's house to fall. A will have no remedy at law against B because A should not have built his house so close to the property line. A may not prevent B from building on his land as he desires. *Id.*

50. *Id.*

51. *See, e.g.*, Morris v. Ostertag, 376 S.W.2d 720, 723 (holding that landowners may not excavate soil when such actions cause injury to adjoining soil in its natural state).

52. 203 S.W. 945 (Tenn. 1918).

53. *Id.* at 946.

54. *Id.* Defendants hired a contractor to perform the job. The contractor was a named defendant, along with the landowners. *Id.*

55. *Id.*

brought suit for damages to their building.[56] The Tennessee Supreme Court ruled in favor of the plaintiffs,[57] holding that the defendants

> had the right to make such alterations or repairs in the part of the wall standing upon the [defendants'] lot as was necessary to erect the proposed building, provided it could be done without weakening or otherwise impairing plaintiffs' use of it, and that in so doing they stood as insurers to plaintiffs against injury to their house growing out of the work.[58]

Thus, when landowners share a party wall, each owes the other a duty not to damage or destroy the wall.[59]

In *Murray v. Patterson*,[60] the Tennessee Court of Appeals did not require the plaintiff to prove negligence to recover under a lateral support cause of action for damage to a party wall.[61] The defendants removed several feet of dirt from below the bottom of the party wall without providing adequate support for the wall.[62] When the party wall collapsed, the rear of the plaintiff's building was destroyed.[63] The plaintiff brought an action against the defendants for damages resulting from the defendants' excavations.[64] The plaintiff alleged that the defendants were negligent in their removal of lateral support, and that the defendants' negligence was the proximate cause of the wall's collapse.[65] The defendants claimed that the wall fell because it was old and damaged.[66] The defendants argued that their actions were not negligent, and that they were merely exercising their right to make improvements to the land.[67] Using principles set forth in *Carroll Blake,* the court ruled that if the wall fell due to

---

56. *Id.*

57. *Id.* Landowners attempted to escape liability by claiming that the company performing the work was an independent contractor whose negligence caused the damages. *Id.* at 947. The court held that both the contractor and the landowners were liable for the damages. *Id.* The court refused to release the landowners from liability, claiming "their duty to prevent injury to plaintiffs' property was absolute." *Id.* Defendants also asserted that the plaintiffs should not recover because the plaintiffs had violated a city ordinance by failing to adequately secure the wall after learning of defendants' excavation plans. *Id.* at 947-48. The court rejected defendants' argument, holding that the city ordinance did not apply to party walls. *Id.*

58. *Id.* at 946.

59. *Id.* at 947 (citing Sanders v. Martin, 70 Tenn. 213, 216 (Tenn. 1879)).

60. 72 S.W.2d 558 (Tenn. Ct. App. 1934).

61. *Id.* at 564.

62. *Id* at 560.

63. *Id.*

64. *Id.*

65. *Id.*

66. *Id.* at 562. Defendants asserted that there was a large crack in the wall visible only to the plaintiff, and because the defendants did not have knowledge of the crack, they were not liable for damages resulting from the crack. *Id.* The court denied the defendants' argument, accepting material evidence that the crack was superficial and did not weaken the wall or cause it to fall. *Id.*

67. *Id.*

the defendants' excavations, then the defendants were liable to the plaintiff without proof of negligence.[68] Based on the evidence, the court concluded that the wall collapsed as a result of the defendants' actions, and the court affirmed the jury verdict for the plaintiff.[69]

After developing fundamental lateral support laws through cases involving party walls, Tennessee courts broadened their application of lateral support principles.[70] In *Morris v. Ostertag*,[71] the Tennessee Court of Appeals held a landowner strictly liable for damage caused by the removal of support from land in its natural state.[72] The defendant purchased property adjacent to the plaintiffs' house in order to build a subdivision.[73] One afternoon, the plaintiffs returned home to discover bulldozers excavating the road next to their house.[74] The defendant's excavators removed portions of the plaintiffs' soil and inadvertently cut into the overflow field of plaintiffs' septic tank.[75] The defendant admitted cutting into the plaintiffs' property, but stated that excavations were made in an effort to conform to Davidson County specifications.[76] In its decision, the court relied on *Levi v. Schwartz*,[77] a Maryland opinion. Drawing from *Levi* and many other sources, the *Morris* court held that the defendant was liable for damages resulting from the removal of naturally necessary lateral support to the plaintiffs' soil.[78] The court

---

68. *Id.* at 562-63.

69. *Id.* at 563.

70. *See generally* Williams v. S. Ry. Co., 396 S.W.2d 98, 100 (Tenn. Ct. App. 1965) (applying lateral support theory to a case involving a "railroad cut"); Morris v. Ostertag, 376 S.W.2d 720, 724 (Tenn. Ct. App. 1963) (applying lateral support theory to a case involving the grading of a subdivision).

71. 376 S.W.2d 720 (Tenn. Ct. App. 1963). For a summary of *Morris*, see Dix W. Noel, *Torts - 1964 Tennessee Survey*, 18 VAND. L. REV. 1263, 1286-87 (1965).

72. *Morris*, 376 S.W.2d at 725-26.

73. *Id.* at 721.

74. *Id.* The defendant hired an independent company to perform the excavations. *Id.* at 721, 722. This independent company was also named as a defendant. *Id.* at 721.

75. *Id.*

76. *Id.* at 722.

77. 95 A.2d 322 (Md. 1952). In *Levi*, the plaintiff landowner gave the defendant rights to use the street leading up to the plaintiff's home as long as the defendant built another road to the property. *Id.* at 325. The plaintiff alleged that while constructing the road, the defendant removed support soil from the property, resulting in damages. *Id.* The defendant claimed that he was not liable for damages, as he was complying with an ordinance establishing the appropriate grade of the road. *Id.* at 325-26. The court held,

> It can reasonably be inferred that [plaintiff] and [defendant] understood that the grades for the streets had to be established by the City Planning Commission. But it can also be reasonably inferred that [defendant] would be liable for loss of lateral support to [plaintiff's] land and for any unlawful invasion of it.

Id. at 326. Recognizing the importance of the principle that adjoining landowners are liable for damage caused by removal of lateral support, the court did not allow the defendant to escape liability based on compliance with a city ordinance. *Id.*

78. *Morris*, 376 S.W.2d at 726. The court held that the defendant developer was liable

emphasized that the right of a landowner to have his soil supported by adjacent property is absolute, and anyone who excavates is liable for the resulting damages without a showing of negligence.[79]

The Tennessee Court of Appeals reconciled Tennessee cases and common law in *Williams v. Southern Railway Co.*[80] In that case, the defendant railroad company made a railroad cut on the land adjoining the plaintiff's property in 1905.[81] Aside from the periodic removal of dirt that covered the railroad tracks, there had been no other alterations to the cut.[82] In 1963, heavy rain caused a landslide into the cut.[83] A portion of the plaintiff's property fell approximately one foot and caused damages to the foundation of a house built on the property.[84] The court reversed the trial court's directed verdict in favor of the defendant and remanded the case.[85] Although it had been nearly 60 years since the defendant had made the cut, the court stated that a reasonable jury could find the damage to be a proximate result of the defendant's removal of lateral support.[86]

In its decision, the *Williams* court stated the well-supported principle of lateral support:

> It is an established principle of law that every owner of land has the right to naturally necessary lateral support from the adjoining soil, and if a landowner removes the soil from his own land so near the land of his neighbor that his neighbor's soil will crumble away under its own weight, he is liable for damages naturally resulting therefrom, including damage to structures on the subsiding land, without the necessity of showing negligence or want of skill on the part of the adjoining owner in making the excavation.[87]

The court further held that "[t]hough the soil falls of its own weight and pressure, the liability for injury to the land attaches; and the fact that the falling of the soil was due to the action of the elements does not constitute a defense."[88] The plaintiff did not need to show that defendant acted negligently to recover.[89] This duty, as emphasized in *Morris*, is absolute.[90] The *Williams*

---

even though an independent company performed the excavations. *Id.* at 725-26. The developer could have anticipated that excavations would cause damage to the plaintiffs' property. *Id.* at 726. Thus, the developer could not escape liability by hiring an independent company to do the work when the damages were foreseeable. *Id.*

79. *Morris*, 376 S.W.2d at 723-24; *see e.g.*, AM. JUR. 2D, *supra* note 22, § 48.
80. 396 S.W.2d 98 (Tenn. Ct. App. 1965).
81. *Id.* at 99-100.
82. *Id.*
83. *Id.*
84. *Id.*
85. *Id.* at 99, 101.
86. *Id.* at 100.
87. *Id.* at 99 (citations omitted).
88. *Id.* at 100-01 (citing AM. JUR. 2D, *supra* note 22, § 43).
89. *Id.* at 99-100.

holding provided continuity to Tennessee's law of lateral support for land in its natural state.

C. *Lateral Support of Land Altered From its Natural State*

Most states recognize that the right of lateral support is limited to land that has not been altered from its natural state.[91] If land has been excavated or filled, adjoining landowners generally have no duty to support the altered property.[92] The adjoining landowner may excavate his property as long as the excavator exercises due care to prevent negligent injury to neighbors.[93] These general principles may seem logical, but the challenge for courts is to determine whether land is in its natural state.

Courts in various jurisdictions have not applied a uniform rule setting forth the rights of lateral support when an individual purchases land that has been altered from its natural state. A Texas court strictly construed the definition of "natural state" in *Carrion v. Singley*.[94] The plaintiff and defendant lived in a neighborhood situated on a hillside, and their respective lots had previously been cut and graded by the developer who sold them the property.[95] To support the plaintiff's lot, which was slightly elevated, the developer built a retaining wall on the defendant's property.[96] The wall began to deteriorate.[97] The plaintiff filed a suit for an injunction, demanding that the defendant repair and maintain the wall.[98] The court held that the defendant did not have to maintain the wall supporting plaintiff's property because both properties had been altered from their natural state.[99]

Other courts, unwilling to support such a harsh outcome, have relaxed the common law rule of lateral support for landowners who purchase property that has been altered from its natural state. The Arkansas Court of Appeals held in *Urosevic v. Hayes*[100] that subsequent purchasers of excavated land owed a

---

90. *Id.* at 100. In contrast, the *Williams* court noted:

> It is where the subsidence of the property is due to the additional weight of buildings constructed thereon and not to the weight of the land in its natural condition that the person damaged must show that the excavator failed to use reasonable care in making and/or maintaining the excavation.

Id. After examining the evidence, the court held that the subsidence of the land was a natural and proximate cause of the cut in the land; thus, the plaintiff did not have to prove negligence. *Id.* at 100-01.

91. *See* AM. JUR. 2D *supra* note 22, § 40; RESTATEMENT, *supra* note 23, § 817 cmt. c.
92. AM. JUR. 2D *supra* note 22, § 42.
93. AM. JUR. 2D *supra* note 22, § 53.
94. 614 S.W.2d 916 (Tex. Civ. App. 1981).
95. *Id.* at 917.
96. *Id.*
97. *Id.*
98. *Id.* at 916.
99. *Id.* at 917.
100. 590 S.W.2d 77 (Ark. Ct. App. 1979).

continuing duty of lateral support to adjoining land even when the current landowners had not made the excavations.[101] The court required both landowners to share the cost of constructing a retaining wall to prevent the plaintiff's land from falling away.[102]

A Pennsylvania state court took a fresh approach to applying the general rule of lateral support in *Durante v. Alba*.[103] The court held that a landowner must continue to provide lateral support to a building on adjoining property, which was in existence when the adjoining landowner purchased the property.[104] Notably, "the purchaser of a portion of a tract of land takes it burdened or benefited by all palpable and manifest qualities annexed thereto, or which appear in his chain of title."[105] Though the court opined that the erection of the building altered the land from its natural state, there remained "an implied duty of lateral support of the building as well as of the soil."[106]

Although the Pennsylvania courts have been hesitant to strictly adhere to the rules of lateral support, the Pennsylvania Supreme Court did apply general lateral support principles in *Hutchinson & Rourke v. Schimmelfeder*[107] by appealing to the parties' morals.[108] In that case, the defendant filled his land so that it would be level with the street, but in doing so, caused the fill to encroach upon the plaintiff's land.[109] The court required the defendant to keep the fill off the plaintiff's land, and suggested that the defendant build a wall on his own land.[110] The court held that the defendant had both a legal and moral obligation to use his land in such a way as to prevent injury to adjoining landowners.[111]

The Pennsylvania courts have broadly defined the term "natural condition," concluding that altered land may revert to its natural condition after the passage

---

101. *Id.* at 79.

102. *Id.* Although courts do not frequently require landowners to share the costs of lateral support, this Arkansas court's self-proclaimed "peculiar" holding may have been influenced, in part, by the possibility that "restoration of the wall will be beneficial to the properties of both" the plaintiff and the defendant. *Id.*

103. 109 A. 796 (Pa. 1920). In *Durante*, a landowner sold the plaintiff a portion of a tract and retained the remainder of the property, which the owner later conveyed to the defendant. *Id.* at 797. The defendant purchased the right to use a wall from the plaintiff. *Id.* The defendant removed support from the party wall, which then fell and destroyed the plaintiff's building. *Id.* The court found that the defendant was liable on two grounds. *Id.* at 798. First, the defendant chose to support the building when he purchased rights to the party wall. *Id.* Second, the defendant was "careless" in removing support from the wall. *Id.*

104. *Id.* at 797.

105. *Id.* (citing Kieffer v. Imhoff, 26 Pa. 438 (1856); Seibert v. Levan, 8 Pa. 383 (1848)).

106. *Id.*

107. 40 Pa. 396 (1861).

108. *Id.* at 399 (holding that the defendant "was, by a moral as well as legal obligation, bound to use his property so as not to injure that of the plaintiff").

109. *Id.* at 396-97.

110. *Id.* at 398.

111. *Id.* at 399.

of time. In *Bradley v. Valicenti*,[112] the plaintiff spread a small amount of fill dirt on his land for the construction of a house.[113] After an adjoining landowner's excavations caused some of the plaintiff's property to fall away, the defendant claimed he owed no duty of lateral support because the land had been altered by fill dirt.[114] The court recognized that the duty of lateral support only applies to land in its natural condition, but determined that "the passage of five years, the resultant use of the land and its exposure to the elements, would make it ground in its natural condition, as viewed by the adjacent land owner and entitled to lateral support by the adjacent property."[115] Five years later, in *Albert v. Wright*,[116] the Pennsylvania Supreme Court narrowed *Bradley*'s broad definition of natural condition. The court held that *significant* amounts of fill dirt altered the plaintiff's land from its natural state and destroyed the adjoining landowner's duty of lateral support.[117] The *Albert* court emphasized that *Bradley* was not intended to create an additional right of lateral support.[118] The court distinguished its holding by emphasizing that in *Bradley*, the amount of fill dirt added to the property was insignificant.[119] In *Albert* the land was not in its natural state because "the amount of fill was considerable, consisting of some 4700 cubic yards."[120] The definition of "natural state" set forth in these Pennsylvania cases is far from being deemed the common law, but it has been supported in other jurisdictions.[121]

In *Bay v. Hein*,[122] the Washington Court of Appeals saw glaring weaknesses in the common law approach to lateral support and, as a result, expanded the duties of adjoining landowners.[123] In *Bay*, the defendant landowner filled and built upon his property.[124] The defendant then sold a portion of the land containing a house to the plaintiff's predecessor, but retained much of the surrounding land.[125] The plaintiff brought suit after the defendant removed naturally necessary lateral support of the plaintiff's

112. .138 A.2d 238 (Pa. Super. Ct. 1958).
113. *Id.* at 239.
114. *Id.*
115. *Id.* at 240.
116. 189 A.2d 753 (Pa. 1963).
117. *Id.* at 755. Landowners had added a considerable amount of fill to the property. *Id.*
118. *Id.*
119. *Id.*
120. *Id.*
121. *See* Biel v. Metro. Life Ins. Co., 304 N.E.2d 449, 451 (Ill. App. Ct. 1973) (holding that because the land had been sold to another landowner and had been exposed to the elements for twelve years, it reverted to its natural condition). *Contra* Bay v. Hein, 515 P.2d 536, 539 (Wash. Ct. App. 1973) (supporting its refusal to apply *Bradley*, the court reasoned, "American courts have generally refused to recognize prescription as a method for changing the content of the duty of lateral support").
122. 515 P.2d 536 (Wash. Ct. App. 1973).
123. *Id.* at 539.
124. *Id.* at 537.
125. *Id.*

property.[126] The court recognized that at common law, the plaintiff would have no remedy because the land had been altered from its natural state.[127] But the court expanded common law rules by holding that the defendant may not excavate his land to an extent that will damage the plaintiff's land.[128] In similar situations, other courts recognize an "implied easement for the reasonably necessary support" of existing buildings or structures.[129] Nonetheless, courts have had difficulty applying these seemingly simple rules of lateral support, a point that is evidenced by these inconsistent holdings.

IV. *XI PROPERTIES, INC. V. RACETRAC PETROLEUM, INC.*: ALIGNING TENNESSEE LAW WITH TRADITIONAL COMMON LAW LATERAL SUPPORT PRINCIPLES

In *XI Properties, Inc. v. RaceTrac Petroleum, Inc.*,[130] the Tennessee Supreme Court addressed whether a landowner owes a duty of "lateral support to adjoining land that has been altered from its natural state"[131] and, if not, whether a landowner may remove support from the altered land, disregarding the interest of the adjoining landowner.[132] Justice Barker wrote the court's unanimous opinion.[133] The court affirmed the trial court's holding that the plaintiff landowner did not owe a duty of lateral support to the defendant because the defendant's land had been altered from its natural state.[134] The court also affirmed the appellate court's holding that the plaintiff may not negligently excavate the property.[135]

A. A Summary of the Duties of Naturally Necessary Lateral Support

In his discussion of lateral support to adjoining land, Justice Barker began by setting forth existing Tennessee law.[136] Justice Barker emphasized the "traditional common law" lateral support principles set forth in *Williams*.[137] Justice Barker set forth these traditional rules of lateral support: (1) an owner of land is entitled to naturally necessary lateral support from adjoining land;[138]

---

126. *Id.*
127. *Id.* at 539.
128. *Id.*
129. JON W. BRUCE & JAMES W. ELY, JR., THE LAW OF EASEMENTS AND LICENSES IN LAND § 4:27 (2001).
130. 151 S.W.3d 443 (Tenn. 2004).
131. *Id.* at 446.
132. *Id.* at 448.
133. *Id.* at 444.
134. *Id.* at 445.
135. *Id.*
136. *Id.* at 447.
137. *Id.*
138. *Id.*

(2) an owner of land who excavates or improves land must perform such undertakings to protect adjoining land in its natural state from falling away;[139] (3) an owner of land may excavate along a boundary line as long as any natural support that is removed is replaced artificially;[140] (4) an owner is strictly liable for damage resulting from the removal of naturally necessary lateral support of property, without proof of negligence.[141] None of these duties of lateral support extend to land that has been altered from its natural state.[142] The property in *XI* had been filled above its natural level to make a parking lot.[143] The plaintiff acquired a slope of land that supported property that had been improved from its natural state.[144] Thus, the court held that the plaintiff owed no absolute duty of lateral support to the defendant's property, as it had been altered from its natural state.[145]

B. Lateral Support of Land Altered From Its Natural State

Justice Barker then turned to the task of differentiating between the rules set forth in *Williams* and the rules of lateral support as applied to land that has been altered or improved.[146] When a landowner fills his land, adjoining landowners do not have a duty to support to the filled portion of the land.[147] In relieving adjoining landowners of their duty of lateral support for filled land, Justice Barker placed the duty on the owners of filled land to prevent it from encroaching on neighboring property.[148] Landowners must keep filled land from falling on adjoining land by building retaining walls, if necessary.[149]

Justice Barker did not completely release landowners from liability if their property adjoins filled property.[150] The court held that "if lateral support is provided to adjoining land containing improvements, such as buildings or other construction, this support cannot be removed in complete disregard of the adjoining landowner's interest."[151] Adjoining landowners owe a duty of reasonable care when improving property that adjoins land with artificial additions.[152] Justice Barker emphasized that the duty to avoid foreseeable harm is based on the goal of preventing landowners from harming neighbors'

139. *Id.*
140. *Id.*
141. *Id.*
142. *Id.* at 448.
143. *Id.* at 449.
144. *Id.*
145. *Id.*
146. *Id.* at 448.
147. *Id.*
148. *Id.* at 449.
149. *Id.*
150. *Id.* at 448.
151. *Id.*
152. *Id.* (citing Law v. Phillips, 68 S.E.2d 452, 457-58 (W. Va. 1952); RESTATEMENT, *supra* note 23, § 819).

property.[153] "The law in this respect attempts to strike a balance between the right of landowners to improve and to use their land, and the duty of landowners to avoid foreseeable harm to improvements on adjoining property."[154]

The court held that while the plaintiff owed no absolute duty to support the defendant's property, the plaintiff did owe a duty of reasonable care to avoid foreseeable damages to the property.[155] The court permitted the plaintiff to remove the sloped embankment, but required the plaintiff to exercise reasonable care.[156] The court held that defendant was responsible for installing a retaining wall, if necessary, to support the fill.[157]

C. *The Effects of XI*

Justice Barker wrote a safe and legally accurate opinion regarding the duties of lateral support to land altered from its natural state. The unanimous decision in *XI* indicates that the issue before the court was resolved many years ago by the courts of other jurisdictions. In his holding, Justice Barker did not examine other possible approaches to lateral support or reasons for rejecting these approaches. Thus *XI*'s greatest weakness is that it fails to address flaws in the common law of lateral support as it relates to today's society.

The *XI* court did not develop a clear definition of what constitutes land in its "natural condition." The defendant asserted that his land reverted to its natural state when it was subdivided and sold to a new owner.[158] Understandably, it would be difficult for a court to determine that a slab of concrete could be considered land in its natural condition. Rather than stating a rationale for rejecting the defendant's assertions, however, the court avoided the issue altogether.

The court's holding in *XI* may affect residential areas. Many of today's neighborhoods are planned subdivisions: a developer purchases large tracts of land, clears and levels the land, builds houses on the land, and sells the land to individual owners. Under the Tennessee Supreme Court's holding in *XI*, homeowners may remove support from neighbors' homes because the land is no longer in its natural condition. As long as landowners exercise a duty of reasonable care, such excavations are legally permitted. The *XI* holding implies

---

153. *Id.* (citing *Law*, 68 S.E.2d at 458).

154. *Id.* Factors courts use to determine whether one is negligent in the removal of lateral support include: "the necessity of the excavation; whether adequate notice was given to adjoining landowners; whether reasonable precautions were taken to prevent harm to adjoining property; whether competent workers were employed; and whether proper instrumentalities were used." *Id*; *see* RESTATEMENT, *supra* note 23, § 819 cmt. e, f.

155. *XI Props., Inc.*, 151 S.W.3d at 449-50.

156. *Id.* at 449.

157 *Id.* The court held, additionally, that the 30 days notice the plaintiffs gave to the defendants regarding the excavation plans was reasonable notice. *Id.*

158. *Id.* at 447.

that land is robbed of its natural state when the developer clears it, regardless of how many years pass after the clearing occurs.

In *Bradley*, the Pennsylvania court avoided such a harsh outcome by ruling that the amount of time that passes after land is altered affects the classification of land in its natural state for purposes of lateral support.[159] Additionally, a Pennsylvania court held in *Durante* that landowners who purchase previously developed property take the tract as it is; those purchasing developed land are entitled to an implied duty of lateral support.[160] The Tennessee Supreme Court did not seek solutions to remedy potential abuse of traditional lateral support laws. In analyzing the *XI* holding, one publication noted that "[i]n some states, [the defendant] would have had an implied support easement, on the theory that the sloped embankment was an open and necessary prior existing use."[161] The Tennessee Supreme Court's holding in *XI* was technically correct, but it failed to take advantage of an opportunity to modernize the existing common law.

V. CONCLUSION

The *XI* holding shields landowners from liability for the improvements adjoining landowners place on their own land. On its face, *XI* is the perfect solution to the issue of lateral support to land that has been altered from its natural state. *XI* closely aligns Tennessee law with the law of a majority of other states and perpetuates common law lateral support doctrines. The Tennessee Supreme Court followed the traditional rules of lateral support by holding that there is no absolute duty of lateral support to land that has been altered from its natural state.[162] But the court did not address difficult underlying issues, such as the definition of land in its "natural condition." *XI* has clarified some of the duties landowners owe to adjoining property that has been altered from its natural state. The technical aspects of adjoining land use have been saved for another day.

MELISSA C. HUNTER

159. Bradley v. Valicenti, 138 A.2d 238, 240 (Pa. Super. Ct. 1958).
160. Durante v. Alba, 109 A. 796, 797 (Pa. 1920).
161. Daniel B. Bogart, *Keeping Current—Property*, PROB. & PROP., May/June 2005, at 32.
162. *XI Props.*, 151 S.W.3d at 449.

SUBSTITUTION

SHAUN P. MARTIN*

Rule 25 is amongst the most obscure and least known of the eighty-nine[1] substantive Federal Rules of Civil Procedure.[2] The Supreme Court has addressed Rule 25 only once.[3] Even then, the Court did so only in a short, unanimous, and seemingly obvious per curiam decision.[4] The lower federal appellate courts are similarly bereft of any substantial discourse regarding Rule 25.[5]

Nonetheless, Rule 25 plays a critical and increasing role in modern civil litigation. Rule 25 governs the substitution of parties after, *inter alia*, the post-filing death or incapacity of a party, or the transfer of their interest in the litigation.[6] Whenever parties merge, die, become incapacitated, or assign their interests, Rule 25 applies and allows the litigation to persist.

Although the federal judiciary increasingly applies Rule 25,[7] this provision has attracted virtually no academic or other intellectual interest. Instead, Rule 25 is universally viewed as containing self-evident and entirely unproblematic principles, both in theory and as applied.

This Article challenges these assumptions. As presently articulated and applied, Rule 25 not only allows substantial circumvention of the complete diversity rule, but does so in a manner that is increasingly misused in the evolving American economic and legal market.[8] Rule 25 also creates numerous

---

* Professor, University of San Diego School of Law. Thanks to Sandra Rierson, Frank Partnoy, David Richter, Peter Stris, Bon Gans, Larry Day, and Judybeth Tropp for their invaluable contributions to this piece.

1. Although the Federal Rules of Civil Procedure begin with Rule 1 and end with Rule 86, there are actually eighty-nine—not eighty-six—different rules. Rules 4.1, 23.1, 23.2, 44.1, 65.1, and 71A add six rules. Rules 74, 75, and 76 were abrogated in 1997.

2. By substantive, one may exclude, *inter alia*, Rule 85, which states: "These rules may be known and cited as the Federal Rules of Civil Procedure."

3. *See* Freeport-McMoRan Inc. v. K N Energy, Inc., 498 U.S. 426 (1991) (per curiam).

4. *See infra* text accompanying notes 24-49 (discussing *Freeport-McMoRan*).

5. Ironically, the Supreme Court's discussion of Rule 25 has attracted a fair amount of judicial attention regarding how to apply the jurisdictional analysis therein to *other* Federal Rules of Civil Procedure; in particular, to Rules 15, 19 and 20. *See infra* text accompanying notes 196-206. By contrast, the Court's direct discussion of Rule 25 has been almost entirely ignored.

6. *See* FED. R. CIV. P. 25(a) (governing substitution of parties upon death); 25(b) (governing substitution of parties upon becoming incompetent); 25(c) (governing substitution or joinder of parties upon a transfer of interest); 26(d) (governing substitution of public officers upon their death or separation from office).

7. *See infra* text accompanying notes 24-41, 64-114 (discussing pertinent Rule 25 cases).

8. *See infra* text accompanying notes 56-57 (The increasing quantity of mergers and acquisitions, heightened importance of intellectual property, escalating significance of business synergies, and accelerating development of markets for the transfer of inchoate legal interests

substantive and jurisdictional problems, particularly given the contemporary breadth of statutory supplemental jurisdiction.[9] This Article identifies these problems and supports a modification of Rule 25 jurisprudence in an attempt to ameliorate them.[10]

I. THE SUBSTANTIVE CONTENT OF RULE 25 SUBSTITUTION

A. The Application of Rule 25

The central components of Rule 25 articulate fairly straightforward and constructive rules regarding the permissible substitution of parties based upon various events that often transpire during the pendency of a federal civil action.[11] Rule 25(a), for example, establishes a procedure that permits the

---

are business practices that might be impacted by a holding that destroys federal diversity jurisdiction because of a Rule 25 transfer of interest.).

9. *See* 28 U.S.C. § 1367 (2000); *infra* Part III.

10. This Article does not attempt to justify or critique the current breadth of federal diversity jurisdiction, a subject regarding which a great deal of academic literature has already been written. *See, e.g.*, Richard Allan, *Demarche or Destruction of the Federal Courts—A Response to Judge Friendly's Analysis of Federal Jurisdiction*, 40 BROOK. L. REV. 637 (1974); Howard C. Bratton, *Diversity Jurisdiction: An Idea Whose Time Has Passed*, 51 IND. L.J. 347 (1976); John P. Frank, *Federal Diversity Jurisdiction—An Opposing View*, 17 S.C. L. REV. 677 (1965); John P. Frank, *For Maintaining Diversity Jurisdiction*, 73 YALE L.J. 7 (1963); Felix Frankfurter, *A Note on Diversity Jurisdiction—In Reply to Professor Yntema*, 79 U. PA. L. REV. 1097 (1931); Carl McGowan, *Federal Jurisdiction: Legislative and Judicial Change*, 28 CASE W. RES. L. REV. 517 (1978); Thomas D. Rowe, Jr., *Abolishing Diversity Jurisdiction: Positive Side Effects and Potential for Further Reforms*, 92 HARV. L. REV. 963 (1979); Thomas D. Rowe, Jr., *Abolishing Diversity Jurisdiction: The Silver Lining*, 66 A.B.A. J. 177 (1980); David L. Shapiro, *Federal Diversity Jurisdiction: A Survey and a Proposal*, 91 HARV. L. REV. 317 (1977); Charles Alan Wright, *Restructuring Federal Jurisdiction: The American Law Institute Proposals*, 26 WASH. & LEE L. REV. 185 (1969); J. Skelly Wright, *The Federal Courts and the Nature and Quality of State Law*, 13 WAYNE L. REV. 317 (1967).

Instead, this Article presumes the existing scope of federal diversity jurisdiction and merely seeks to rationalize the dictates of Rule 25.

11. Rule 25 has a plethora of state law analogues. *See, e.g.*, Cooke v. Walter Kidde & Co., 394 N.E.2d 968, 970 (Mass. App. Ct. 1979) (applying analogous federal common law to substitution issues arising under Massachusetts Rule of Civil Procedure 25, a rule substantively identical to federal Rule 25). Rule 25 has also occasionally been applied to federal civil cases on appeal notwithstanding the general inapplicability of the Federal Rules of Civil Procedure. *See, e.g.*, Total Containment, Inc. v. Intelpro Corp., Nos. 99-1059, 99-1060, 1999 WL 717946, at **3 (Fed. Cir. Sept. 15, 1999) (relying on Rule 25 to authorize the continued prosecution of an appeal notwithstanding the formal lack of standing by the current appellant as a result of appellant's post-judgment transfer of interest). The propriety of these latter holdings has previously been assumed. Yet, they are at best questionable. *Compare id.* (applying Rule 25 on appeal because "there is no direct provision on this subject in the Federal Rules of Appellate

successors or representatives of a deceased party to be substituted for the decedent, as long as the underlying substantive claim is not extinguished by death.[12] Rule 25(b) performs analogous functions with respect to parties who become incompetent during the pendency of an action.[13] Rule 25(d) similarly addresses the substitution of public officers who "die[], resign[], or otherwise cease[] to hold office."[14]

Procedure") *with* FED. R. APP. P. 43 (directly addressing the substitution of parties on appeal and expressly governing appellate substitution "for any reason other than death").

Nonetheless, although the conclusions of this Article may have consequences for both state court and appellate litigation, this Article focuses centrally on the direct application of Rule 25 by federal district courts with only tangential comments regarding related Rules and alternative fora. This Article leaves to others an assessment of the analogous implications of these principles on state court, appellate, and post-judgment proceedings. *Cf.* Ciccone v. Sec'y of Dep't of Health & Human Servs., 861 F.2d 14, 15 n.1 (2d Cir. 1988) (stating that the failure to file a motion to substitute pursuant to FED. R. APP. P. 43 does not divest the Court of Appeals of the power to adjudicate the appeal); Otis Clapp & Son, Inc. v. Filmore Vitamin Co., 754 F.2d 738, 743 (7th Cir. 1985) (holding that denial of Rule 25(c) motion to substitute does not preclude subsequent action against proposed defendant to enforce judgment); Air Line Pilots Ass'n Int'l v. Tex. Int'l Airlines, Inc., 567 F. Supp. 78, 81 (S.D. Tex. 1983) (permitting post-judgment substitution).

12. *See* FED. R. CIV. P. 25(a)(1) ("If a party dies and the claim is not thereby extinguished, the court may order substitution of the proper parties."). Rule 25(a) not only authorizes substitution, but also establishes the procedures through which substitution must be made. *See id.* ("The motion for substitution may be made by any party or by the successors or representatives of the deceased party and, together with the notice of hearing, shall be served on the parties as provided in Rule 5 and upon persons not parties in the manner provided in Rule 4 for the service of a summons, and may be served in any judicial district.").

Similarly, Rule 25(a) provides for the dismissal of a deceased, nonsubstituted party, as well as for the continuation of civil actions in which only a portion of the claims therein have been abated by the party's death. *See* FED. R. CIV. P. 25(a)(1) ("Unless the motion for substitution is made not later than 90 days after the death is suggested upon the record by service of a statement of the fact of the death as provided herein for the service of the motion, the action shall be dismissed as to the deceased party."); *id.* 25(a)(2) ("In the event of the death of one or more of the plaintiffs or of one or more of the defendants in an action in which the right sought to be enforced survives only to the surviving plaintiffs or only against the surviving defendants, the action does not abate. The death shall be suggested upon the record and the action shall proceed in favor of or against the surviving parties.").

13. *See* FED. R. CIV. P. 25(b) ("If a party becomes incompetent, the court upon motion served as provided in subdivision (a) of this rule may allow the action to be continued by or against the party's representative.").

14. FED. R. CIV. P. 25(d)(1) ("When a public officer is a party to an action in his official capacity and during its pendency dies, resigns, or otherwise ceases to hold office, the action does not abate and the officer's successor is automatically substituted as a party. Proceedings following the substitution shall be in the name of the substituted party, but any misnomer not affecting the substantial rights of the parties shall be disregarded. An order of substitution may be entered at any time, but the omission to enter such an order shall not affect the substitution."); *id.* 25(d)(2) ("A public officer who sues or is sued in an official capacity may be

The content of Rules 25(a), (b), and (d) do not have any substantial incentive effects. A litigant would be unlikely to choose to die, become incompetent, or leave public office merely to obtain the procedural advantages that could arise from the permissible substitution of a representative or successor litigant. The content of these Rules is nonetheless neither self-evident nor entirely well-founded. Rule 25(d) in particular contains a uniquely substantive abatement rule[15] that may constitute an impermissible violation of the Rules Enabling Act.[16] Moreover, even though Rules 25(a), (b), and (d) create no incentive problems, they may engender substantial jurisdictional problems in those cases in which they apply.[17] Nevertheless, because the underlying circumstances that give rise to the applicability of Rule 25(a), (b) and (d) are only marginally capable of advantageous manipulation by the parties, the systemic misuse of these provisions is unlikely.

By contrast, Rule 25(c) authorizes the substitution of parties in contexts that may be deliberately controlled by litigants, and hence may be subject to expansive abuse. Rule 25(c), entitled "Transfer of Interest" provides: "In the case of a transfer of interest, the action may be continued by or against the original party, unless the court upon motion directs the person to whom the interest is transferred to be substituted in the action or joined with the original party."[18] Substitution under Rule 25(c) is classically employed when an

---

described as a party by the officer's official title rather than by name; but the court may require the officer's name to be added.").

15. Rule 25(a)-(c) does not contain any substantive content. In fact, several provisions expressly decline to articulate any federal abatement rule. *See* FED. R. CIV. P. 25(a)(1) (applying only "[i]f a party dies and the claim is not thereby extinguished"); 25(a)(2) (applying only to those circumstances "in which the right sought to be enforced survives only to the surviving plaintiffs or only against the surviving defendants"). Thus, Rule 25(a)-(c) uniformly incorporates the underlying substantive law, whether state or federal, regarding whether any claim for relief abates. *See, e.g.*, *In re* Nuckolls v. McClay, No. 88-1069-C, 1990 WL 66605, at *2 (D. Kan. Apr. 26, 1990); Knauer v. Johns-Manville Corp., 638 F. Supp. 1369, 1387 (D. Md. 1986); Asklar v. Honeywell, Inc., 95 F.R.D. 419, 422 (D. Conn. 1982); Fletcher v. Grinnel Bros., 64 F. Supp. 778, 779 (E.D. Mich. 1946).

By contrast, Rule 25(d) expressly provides that "[w]hen a public officer is a party to an action in his official capacity and during its pendency dies, resigns, or otherwise ceases to hold office, the action does not abate and the officer's successor is automatically substituted as a party." FED. R. CIV. P. 25(d)(1). This provision facially establishes a substantive federal rule of abatement. Moreover, these provisions are equally applicable not only to federal and state claims, but also to federal and state officers.

16. *See* 28 U.S.C. § 2072(a)-(b) (2000) ("The Supreme Court shall have the power to prescribe general rules of practice and procedure and rules of evidence for cases in the United States district courts (including proceedings before magistrates thereof) and courts of appeals. . . . Such rules shall not abridge, enlarge or modify any substantive right.").

17. *See infra* Part III.

18. FED. R. CIV. P. 25(c) (providing that the service of a motion to substitute under Rule 25(c) is to be made pursuant to the provisions articulated in Rule 25(a)).

individual litigant assigns his rights or obligations to a third party[19] or when a business organization either merges with or dissolves into a distinct legal entity.[20] Rule 25(c) attempts to advance the same fundamental objective as the other portions of Rule 25, which allow the continued and efficacious resolution of an existing civil action notwithstanding the occurrence of post-filing events that might potentially transform the nature or status of a particular litigant as an enduring real party in interest.[21]

Rule 25 undoubtedly accomplishes this central objective. The substantial procedural flexibility granted to federal district courts by Rule 25 typically enables the rapid and effective substitution of a successor for an existing party, and facilitates the continued prosecution and defense of the action. The core function of Rule 25 is not controversial in theory or systemic application.

Nonetheless, a bevy of actual and potential jurisdictional problems arise as a direct consequence of the laudable objectives advanced by Rule 25. These problems are particularly noticeable when the substitution results from the potentially tactical decision by a party to transfer its interest in the underlying

---

19. *See, e.g.*, Fischer Bros. Aviation, Inc. v. NWA, Inc., 117 F.R.D. 144, 147 (D. Minn. 1987) (permitting former stockholders of dissolved corporate party plaintiff to substitute for original plaintiff under Rule 25(c) in light of corporation's post-filing assignment of the antitrust claims at issue); Gen. Battery Corp. v. Globe-Union, Inc., 100 F.R.D. 258, 264 (D. Del. 1982) (allowing substitution when ownership transferred from parent to subsidiary); Vandenbark v. Busiek, 1 F.R.D. 366, 366 (E.D. Ill. 1940) (allowing joinder as party plaintiff under Rule 25(c) of individual to whom all interests in the existing litigation were assigned); *see also* 7C CHARLES ALAN WRIGHT, ARTHUR R. MILLER, & MARY KAY KANE, FEDERAL PRACTICE AND PROCEDURE § 1958 (2d ed. 1986) ("[Rule 25] applies to ordinary transfers and assignments, as well as to corporate mergers.").

20. *See, e.g.*, Virgo v. Riviera Beach Assocs., Ltd., 30 F.3d 1350, 1358 (11th Cir. 1994) (permitting plaintiff in sexual harassment suit to substitute a third party who purchased all of the assets of an existing party defendant); Hirsch v. Bruchhausen, 284 F.2d 783, 786 (2nd Cir. 1960) (permitting substitution of partnership for corporate defendant that was dissolved during pendency of the action); FDIC v. Wrapwell Corp., 922 F. Supp. 913, 917 (S.D.N.Y. 1996) (allowing substitution of the FDIC as the plaintiff in an action originally filed by the First New York Bank for Business upon the post-filing appointment of the FDIC as the Bank's receiver); Wainwright v. Kraftco Corp., 58 F.R.D. 9, 13 (N.D. Ga. 1973) (allowing addition under Rule 25(c) of corporate entity who purchased all assets of existing party defendant regardless of whether this activity constituted a valid "merger" under Georgia law); *see also* Samuel Adams & James J. Arguin, *Parties*, *in* 1 BUSINESS AND COMMERCIAL LITIGATION IN FEDERAL COURTS § 14.7(d)(1) (Robert L. Haig ed., 1998) ("Besides the ordinary assignment cases, other common FRCP [sic] 25(c) situations include: resignation of a trustee and appointment of a new trustee, appointment of a trustee in bankruptcy for one of the parties; and a corporate merger involving one of the parties.").

21. *Cf.* FED. R. CIV. P. 17(a) ("Every action shall be prosecuted in the name of the real party in interest. . . . No action shall be dismissed on the ground that it is not prosecuted in the name of the real party in interest until a reasonable time has been allowed after objection for ratification of commencement of the action by, or joinder or substitution of, the real party in interest; and such ratification, joinder, or substitution shall have the same effect as if the action had been commenced in the name of the real party in interest.").

litigation and seek to join the transferee pursuant to Rule 25(c). Joinder through Rule 25 in such cases could permit deliberate jurisdictional manipulation.

The ability to utilize Rule 25 in such cases as a means to invoke or defeat the exercise of federal subject matter jurisdiction has not gone unnoticed by able counsel.[22] Although the Supreme Court has addressed the basic issues that may arise from such conduct,[23] Rule 25 continues to be subject to a vast amount of manipulation and abuse, particularly as a means of circumventing the core jurisdictional requirement of complete diversity between the parties.

B. Rule 25 in the Supreme Court

The Supreme Court's opinion in *Freeport-McMoRan Inc. v. K N Energy, Inc.*[24] is the leading and only Supreme Court decision on the jurisprudential dictates of Rule 25. Furthermore, *Freeport-McMoRan* also involves a factual setting that illuminates at least one manner in which Rule 25 might potentially be abused in an attempt to thwart the complete diversity rule.

Freeport-McMoRan arose from a breach of contract action jointly filed by McMoRan Oil and Gas Company ("McMoRan") and its parent company, Freeport-McMoRan ("Freeport"), against K N Energy, Inc. ("KN").[25] Freeport and McMoRan alleged that KN failed to pay agreed-upon prices for natural gas, and sought both monetary damages and prospective declaratory relief.[26] Federal diversity jurisdiction clearly existed at the outset of the litigation: "Freeport and McMoRan were [each] Delaware corporations with their principal places of business in [New York and] Louisiana. KN was . . . a Kansas corporation with its principal place of business in Colorado."[27]

---

22. *See, e.g.*, discussion *infra* Part III.B. The functional significance of Rule 25 in such settings is facially party-neutral. Plaintiffs and defendants alike have attempted to employ Rule 25 as a means of modifying the presence of federal jurisdiction. Rule 25 has similarly been employed both in an attempt to defeat such jurisdiction (by moving for dismissal based upon the substitution of a nondiverse party) as well as in an attempt to expand it (through the joinder of additional supplemental claims). *See id.* The sole common character of the existing authorities is the tactical utilization of Rule 25 by whichever party perceives maximal advantage from the expansion, limitation, or destruction of federal diversity jurisdiction.

23. *See infra* Part II.

24. 498 U.S. 426 (1991) (per curiam).

25. *Id.* at 427.

26. *Id.*

27. *Id.*; *see also* 28 U.S.C. § 1332 (2000) (authorizing the exercise of federal subject matter jurisdiction based on diversity over such actions). It is perhaps fortunate that the Supreme Court's decision in *Freeport-McMoRan* was articulated in a per curiam opinion with an unknown author. To the undying shame of the relevant Justice and his or her law clerk, the opinion contains a glaring and previously overlooked factual error. The principal place of business of Freeport was not, as articulated by the Court, Louisiana, but was actually New York. *Compare Freeport-McMoRan*, 498 U.S. at 427 (stating that Freeport's principal place of

There would be no continuing jurisprudential significance to the contractual dispute at issue in *Freeport-McMoRan* had McMoRan simply continued its existing business activities and the underlying litigation.[28] However, after the federal lawsuit was filed, McMoRan decided—for commercial reasons wholly unrelated to the action—to transfer its interest in the natural gas contract with KN to a limited partnership named FMP Operating Company ("FMPO").[29] In light of this post-filing transfer, shortly before trial McMoRan requested that the district court allow substitution of FMPO as a party plaintiff pursuant to Rule 25(c).[30] The district court granted the Rule 25(c) motion, added FMPO to the action, and after a two-day bench trial entered judgment in favor of the plaintiffs for $1,602,712.51.[31]

On appeal, however, the Tenth Circuit reversed.[32] The court concluded that the Rule 25 substitution of FMPO had divested the district court of subject matter jurisdiction because both KN ("a Kansas corporation with its principal place of business in Colorado") and FMPO (a limited partnership with individual members who were citizens of Colorado and Kansas) were deemed

---

business was Louisiana) with McMoRan Oil & Gas Co. v. KN Energy, Inc., 907 F.2d 1022, 1023 (10th Cir. 1990) (describing Freeport's principal place of business as New York).

28. In retrospect, McMoRan would perhaps have been better served to have done precisely that. The voluntary transfer that gave rise to the Rule 25 substitution in *Freeport-McMoRan* compelled McMoRan to engage in expensive appellate litigation and to endure substantial delay before it could seek to collect the judgment against KN.

29. *See Freeport-McMoRan*, 498 U.S. at 427.

30. *Id.* Rule 25 is generally described as permitting the "substitution" of an existing party by a distinct entity. *See* FED. R. CIV. P. 25(a) ("[T]he court may order substitution of the proper parties."); FED. R. CIV. P. 25(b) ("[T]he court . . . may allow the action to be continued by or against the party's representative."); FED. R. CIV. P. 25(d) ("[T]he officer's successor is automatically substituted as a party.").

Unlike the other provisions of Rule 25, subsection (c) alternately authorizes the joinder of the successor entity as opposed to a substitution. *See* FED. R. CIV. P. 25(c) ("In case of any transfer of interest, . . . the court upon motion [may] direct[] the person to whom the interest is transferred to be substituted in the action *or joined with the original party*.") (emphasis added). McMoRan moved to add FMPO as an additional party plaintiff pursuant to Rule 25(c) rather than to have FMPO take its place. *See Freeport-McMoRan*, 498 U.S. at 427; *see also McMoRan*, 907 F.2d at 1025 (noting, *inter alia*, this distinction in an attempt to justify its result). According to the Supreme Court, however, there is no jurisdictional significance to the Rule 25 joinder of a party as opposed to the exchange of an original party for another. *See Freeport-McMoRan*, 498 U.S. at 427.

Accordingly, the jurisdictional consequences of a Rule 25 substitution are the same regardless of whether the new party enters the action as a replacement or as an additional party. *But see infra* text accompanying notes 196-225 (discussing subsequent jurisdictional issues, particularly as applied to removal cases, that may be affected by the manner in which the successor party is added to the action pursuant to Rule 25(c), e.g., through substitution versus joinder).

31. *See McMoRan*, 907 F.2d at 1023.

32. *Id.* at 1025.

to be citizens of Kansas and Colorado, thereby destroying complete diversity.[33] The Tenth Circuit held that Rule 25 did not authorize either the abrogation of the complete diversity rule or the exercise of ancillary jurisdiction[34] over FMPO's claims against KN, and accordingly instructed the district court to dismiss the action for lack of jurisdiction.[35]

The Supreme Court, in a unanimous per curiam decision,[36] reversed the Tenth Circuit, holding that the substitution of a nondiverse party under Rule 25 did not deprive the court of federal subject matter jurisdiction.[37] The Court noted that it "ha[d] consistently held that if jurisdiction exists at the time an action is commenced, such jurisdiction may not be divested by subsequent events."[38] Because the plaintiffs and defendant were completely diverse at the time the breach of contract action was initially filed, the subsequent Rule 25(c) substitution of FMPO created by McMoRan's post-filing transfer of interest did

---

33. *Id.* at 1023-25. After the decision of the district court but before the appeal was resolved by the Tenth Circuit, the Supreme Court decided Carden v. Arkoma Assocs., 494 U.S. 185, 195-96 (1990), which held that limited partnerships were deemed to be citizens of every State in which one of their constituent general and limited partners resides.

Judge Moore was none-too-subtle about placing the blame for the resulting jurisdictional dismissal—and attendant waste of judicial resources—on the Supreme Court's intervening (and fairly controversial) 5-4 decision in *Carden*. *See McMoRan*, 907 F.2d at 1022-23 ("Had the Supreme Court recently not decided to 'adhere to our oft-repeated rule that diversity jurisdiction in a suit by or against the entity depends on the citizenship of "all the members,"' this case might well have continued on its journey in federal court, successfully slipping past Scyllaea. Instead, snared on her rocky shore like the misfortuned seafarer, the action is wrest from our jurisdiction and must be dismissed." (quoting *Carden*, 494 U.S. at 195)) (citation omitted).

34. The Tenth Circuit decided *McMoRan* five months prior to the effective date of 28 U.S.C. § 1367, and accordingly applied the common law doctrine of ancillary and pendant jurisdiction rather than the statutory principles of supplemental jurisdiction. *See McMoRan*, 907 F.2d at 1025. Nevertheless, as applied to Rule 25 there is no difference between the operations of these competing jurisprudential regimes, at least according to the analysis articulated by the Supreme Court. *See Freeport-McMoRan*, 498 U.S. at 428 (articulating a holding equally applicable to both); *see also infra* text accompanying notes 265-82 (discussing issues that arise when statutory supplemental jurisdiction is applied to Rule 25).

35. *See McMoRan*, 907 F.2d at 1025.

36. Justice Souter, who was confirmed contemporaneously with the filing of the petition for certiorari in *Freeport-McMoRan*, did not participate in the case. *Freeport-McMoRan*, 468 U.S. at 429.

37. *Id.* The Court's per curiam decision in *Freeport-McMoRan* was tellingly rendered within three months of the filing of the petition for certiorari, a fact that both exemplifies the perceived simplicity of the case as well as perhaps partially explains the Court's pervasive carelessness in the opinion. *See supra* note 27; *infra* notes 38, 40.

38. *Freeport-McMoRan*, 498 U.S. at 428; *see also* Wichita R.R. & Light Co. v. Pub. Utils. Comm'n of Kan., 260 U.S. 48 (1922); Clarke v. Mathewson, 37 U.S. 164 (1838); Mollan v. Torrance, 22 U.S. 537 (1824). Note that the Court's citation to *Wichita Railroad* is inconsistent with Bluebook form, which is another entirely minor but nonetheless embarrassing error.

not compel a dismissal for lack of jurisdiction.[39] The Court held: "Diversity jurisdiction, once established, is not defeated by the addition of a nondiverse party to the action."[40] The Court concluded on these grounds that substitution pursuant to Rule 25 did not alter the preexisting presence of federal subject matter jurisdiction over the existing action, a core holding that persists to this day and greatly facilitates the practical availability of post-litigation transfers of interest.[41]

II. THE BREADTH AND (PARTIAL) VALIDITY OF CONTEMPORARY RULE 25 JURISPRUDENCE

Despite the plethora of conceptual and other flaws in the opinion, the Supreme Court's central holding in *Freeport-McMoRan* is entirely correct and rightly noncontroversial.[42] *Freeport-McMoRan* not only achieves an

---

39. *Freeport-McMoRan*, 468 U.S. at 428.

40. *Id.* This statement (all humor aside) is indicative of the sloppy manner in which the Court wrote its per curiam opinion in *Freeport-McMoRan*. Moreover, unlike the factual, stylistic, and other errors previously discussed, this doctrinal error actually made a difference. The Court was clearly correct that "[d]iversity jurisdiction, once established, is not defeated by the addition of a nondiverse party to the action," *but only if that addition is accomplished pursuant to Rule 25. See* discussion *infra* Part III.A. Omission of this crucial caveat transformed this statement from an entirely accurate maxim to a doctrinal principle that is flatly wrong in virtually every one of its possible manifestations. In fact, it is accurate only as applied to Rule 25. *See, e.g.*, Owen Equip. & Erection Co. v. Kroger, 437 U.S. 365, 367-69, 377 (1978) (holding no diversity jurisdiction over claim by plaintiff against additional nondiverse defendant added pursuant to Rule 14); Travelers Indem. Co. v. Dingwell, 884 F.2d 629, 636 (1st Cir. 1989) (holding no diversity jurisdiction over nondiverse parties added pursuant to Rule 19); Lewis v. Lewis, 358 F.2d 495, 502 (9th Cir. 1966) (holding no diversity jurisdiction when a nondiverse indispensable party is joined); *see also* 28 U.S.C. § 1367(b) (2000) (no supplemental jurisdiction in actions based solely on diversity over claims by plaintiff against persons made parties under Rule 14, 19, 20 or 24 or over claims by additional party plaintiffs joined pursuant to Rules 19 or 24); *infra* Part III.A (discussing at length the widespread and troubling disputes that arose in lower federal courts as a result of this sweeping jurisdictional statement in *Freeport-McMoRan*).

Particularly in light of the brevity of the decision, *Freeport-McMoRan* may have the dubious badge of distinction as the opinion with the highest error-to-length ratio in the history of the Court.

41. *Freeport-McMoRan*, 468 U.S. at 428-29; *see also* discussion *infra* Part III.C (discussing and critically examining the extension of the central holding in *Freeport-McMoRan* as applied to Rule 25 and non-Rule 25 settings).

42. The facial ease of the resulting adjudication is evidenced not only by the wholesale absence of academic commentary following the issuance of the opinion but also by the Court's willingness to summarily reverse the case in a unanimous per curiam opinion without oral argument or a full briefing on the merits by the respondent. *See generally* Boyce F. Martin, Jr., *In Defense of Unpublished Opinions*, 60 OHIO ST. L.J. 177 (1999) (arguing that the practical utility of unpublished opinions outweighs any criticisms). *But cf.* Note, *Supreme Court Per Curiam Practice: A Critique*, 69 HARV. L. REV. 707 (1956) (acknowledging the administrative

appropriate result, but does so on the basis of several principles that legitimately substantiate the Court's conclusions regarding the jurisdictional consequences of substitution under Rule 25. Before an assessment of the potential deficiencies in the Court's analysis in *Freeport-McMoRan*, it is accordingly vital to identify and discuss the core aspects of the opinion that are demonstrably right.

A. The Doctrine

First, as a doctrinal matter, the Court appropriately focused on the circumstances that initially existed at the time the civil action was filed in assessing the impact of a Rule 25 substitution on the persistence of federal subject matter jurisdiction. Although diversity is generally ascertained at the outset of the lawsuit,[43] the subsequent joinder of additional nondiverse parties may terminate this preexisting jurisdictional foundation. For example, when an entirely separate and nondiverse defendant is joined to the action, diversity jurisdiction is clearly destroyed.[44]

The transfer of an existing interest under Rule 25 from a current party to another is not, however, akin to the joinder of an additional unrelated party. Such a change is instead more closely analogous to a post-filing change in the plaintiff's residence—an act that does not eviscerate diversity jurisdiction[45]—than the assertion of additional claims against supplemental defendants. Accordingly, the Supreme Court in *Freeport-McMoRan* was correct to largely apply the "diversity-at-the-time-of-filing"[46] principle to Rule 25 rather than an alternate approach.

---

utility of current *per curiam* practice while criticizing its tendency to engender confusion); William Glaberson, *Caseload Forcing Two-Level System for U.S. Appeals: More One-Word Rulings*, N.Y. TIMES, Mar. 14, 1999, at A1 ("[T]he proliferation of one-word decisions and other abbreviated procedures is shortchanging litigants in the appellate process by limiting the consideration given to their cases by appeals judges.").

43. *See, e.g.*, Linardos v. Fortuna, 157 F.3d 945, 947 (2nd Cir. 1998); Stock West Corp. v. Taylor, 964 F.2d 912, 917 (9th Cir. 1992).

44. *See e.g.* cases cited *supra* note 40.

45. *See, e.g.*, Chapman v. Currie Motors, Inc., 65 F.3d 78, 81 (7th Cir. 1995) (holding that diversity of citizenship, once established at the outset of trial, is unaffected by later change in residence); Bank One, Texas, N.A. v. Montle, 964 F.2d 48, 49 (1st Cir. 1992) (holding lack of post-filing diversity of citizenship irrelevant to continuation of preexisting subject matter jurisdiction); *see also* Peterson v. Allcity Ins. Co., 472 F.2d 71, 74 (2nd Cir. 1972) (holding that plaintiff may validly create diversity jurisdiction by a bona fide change of domicile *before* filing suit even if the motive for such a change is purely jurisdictional); *Lewis*, 358 F.2d at 502 (holding that diversity jurisdiction persisted notwithstanding post-filing change in residence of plaintiff to same state as defendant).

46. *Freeport-McMoRan*, 498 U.S. at 428 ("We have consistently held that if jurisdiction exists at the time the action is commenced, such jurisdiction may not be divested by subsequent events.").

The Court's conclusion is most self-evident in the context of substitutions under Rules 25(a) (death), 25(b) (incompetence), and 25(d) (removal from public office), but is equally true for the type of transfer of interest pursuant to Rule 25(c) at issue in *Freeport-McMoRan*. The replacement of Original Party *X* by Successor Party *Y* upon the post-filing death, incompetence, irrelevance, or official dethronement of Party *X* does not, as a general matter, substantially alter the core assumptions upon which the assertion of federal subject matter jurisdiction was initially based. The efficiencies made available by the presence of a continuing federal forum persist in such cases despite the post-filing transfer of interest, just as they do when a party changes its legal residence during the pendency of an action. Similarly, unlike the joinder of an additional nondiverse party, a substitution under Rule 25 typically affects the merits insubstantially and consists only of a technical alteration in the nature of the formal real party in interest. The same can also be said for a party's post-filing change of residence, and in both cases, the need for the initially proper federal forum persists. The typical jurisdictional significance of a Rule 25 substitution is thus more properly viewed as a post-filing transformation in the legal status of an existing party—an event that typically does not oust the court of jurisdiction—rather than as an impermissible means of joinder.[47]

Accordingly, the Court in *Freeport-McMoRan* appropriately focused on the conditions at the outset of the litigation as central to the continuing ability of a federal court to exercise subject matter jurisdiction. The Court's doctrinal approach does not suggest that post-filing events are entirely irrelevant to the presence of jurisdiction, and such is indisputably not the case.[48] Nonetheless, the Court properly explored the jurisdictional implications of a substitution pursuant to Rule 25 with a persistent emphasis on the conditions that existed at the time the action was initially filed.[49] The Court's analytical assessment in *Freeport-McMoRan* correctly viewed Rule 25 substitution as principally extraneous to the continuing ability of the federal judiciary to assert jurisdiction over an action that was properly filed at the outset in a federal forum.

B. The Reason

The central jurisdictional assumptions of *Freeport-McMoRan* are correct not only as a doctrinal matter, but are supported by the policy reasons upon which the Court relied in reaffirming the equitable character of its holding. In *McMoRan*, the Tenth Circuit recognized that "considerations of efficiency and judicial economy" somewhat militated against a finding that an otherwise permissible substitution under Rule 25 might nonetheless divest the district

47. *But see* discussion *infra* Part II.D (discussing exceptions to the propriety of this general truth).

48. *See supra* note 40; *see also infra* text accompanying notes 184-95 (discussing the proper rule applicable to post-filing events arising in non-Rule 25 settings).

49. *Cf.* discussion *infra* Part III.A (applying this focus on initial jurisdictional conditions even with respect to situations in which Rule 25 substitution is arguably improper).

court of jurisdiction,[50] but held that such concerns were outweighed by the doctrinal consequences that necessarily attended the entry of a nondiverse party into the action.[51] By contrast, the Supreme Court rejected the Tenth Circuit's holding in large part based upon the legal realist[52] belief that the articulation of such a rule "could well have the effect of deterring normal business transactions during the pendency of what might be a lengthy litigation."[53] Moreover, deleterious economic consequences of this type would be unnecessary because, as the Court concluded, the "rule is not in any way required to accomplish the purposes of diversity jurisdiction."[54] On this policy basis, the Court reversed the judgment below.[55]

The *Freeport-McMoRan* Court was undoubtedly correct—both systemically and as applied to the particular case before it—about the potentially adverse results that might arise from a decision to attach dispositive jurisdictional significance to a post-litigation transfer of interest pursuant to Rule 25.[56] Litigants in the modern era increasingly merge, divest themselves of particular assets, or engage in other transactions intended to amplify economic synergies or diminish inefficiencies.[57] These operations typically have little to do with the underlying litigation, but may nonetheless practically, tactically, or legally compel substitution of one or more of the litigants. If substitution

---

50. McMoRan Oil & Gas Co. v. KN Energy, Inc., 907 F.2d 1022, 1025 (10th Cir. 1990).

51. *See id.*

52. *See generally* JOHN HART ELY, DEMOCRACY AND DISTRUST: A THEORY OF JUDICIAL REVIEW 43-72 (1980) (defining legal realism in relation to interpretivism); RICHARD A. POSNER, OVERCOMING LAW 2-3, 20, 393 (1995) (viewing legal realism as a "much-overblown movement" which improperly relies on nonlegal reasoning and lacks rigorous academic inquiry); FREDERICK SCHAUER, PLAYING BY THE RULES 191-96 (1991) (discussing challenges to the plausibility of legal realism); Brian Leiter, *Rethinking Legal Realism: Toward a Naturalized Jurisprudence*, 76 TEX. L. REV. 267 (1997) (clarifying some misrepresentations and misunderstandings about legal realism); Jon O. Newman, *Between Legal Realism and Neutral Principles: The Legitimacy of Institutional Values*, 72 CAL. L. REV. 200 (1984) (exploring legal realism from a judge's perspective). Legal realists would find both interesting and not entirely a coincidence that the business-related concerns persuasive to the increasingly conservative Supreme Court in *Freeport-McMoRan* are generally distinct from those upon which, for example, the Warren Court typically relied.

53. Freeport-McMoRan Inc. v. K N Energy, Inc., 498 U.S. 426, 428-29 (1991) (per curiam).

54. *Id.* at 429.

55. *Id.*

56. Although the jurisdictional rule articulated in *Freeport-McMoRan* facially governs any type of substitution pursuant to Rule 25, the policy-based justification articulated by the Court in fact applies only to voluntary transfers of interest pursuant to Rule 25(c), as parties rarely die, become incompetent, or resign office for reasons that might be altered by the jurisdictional implications of such an event.

57. *See, e.g.*, Joseph A. Flom, *Mergers & Acquisitions: The Decade in Review*, 54 U. MIAMI L. REV. 753, 774-75 (2000) (discussing numerous factors that contributed to the contemporary increase in economic merger, acquisition, and divestment activities).

pursuant to Rule 25 potentially divested the district court of jurisdiction, beneficial business transactions might be deterred, thereby resulting in suboptimal economic arrangements. The damaging practical consequences of such a rule would often also affect the underlying litigation, and could produce inefficacious representation as well as impose substantial transaction costs on the parties and the judiciary. Finally, in particular cases, a holding that reposed dispositive jurisdictional significance upon a Rule 25 substitution would be fundamentally unfair, especially when the substitution occurred after entry of an otherwise binding judgment and thereby stripped this finding of validity and the victor at trial of his legitimate fruits.

Accordingly, as a systemic matter, the Supreme Court in *Freeport-McMoRan* properly recognized the substantial harm that might arise from a perpetuation of the Tenth Circuit's view regarding the impact of Rule 25 substitution on the persistence of federal diversity jurisdiction.[58] The Supreme Court's policy-based justifications for the holding in *Freeport-McMoRan* were not merely theoretical but were also demonstrably exemplified by the underlying facts of the particular case before the Court. The transfer of interest at issue in *Freeport-McMoRan* indisputably transpired for business reasons entirely unrelated to the litigation.[59] Moreover, this transaction occurred after more than two years of pretrial discovery and mere weeks before a trial on the merits was scheduled to begin.[60] Indeed, it was partially on this basis that the district court refused to dismiss the action for lack of subject matter jurisdiction, "noting that with trial just two weeks off, 'at a late point like this, there is sufficient ancillary jurisdiction that the court should proceed.'"[61] Furthermore, by the time *Freeport-McMoRan* reached the Supreme Court, there had already been a trial on the merits, where the original and substituted plaintiffs (McMoRan and FMPO, respectively) had prevailed and obtained a substantial judgment.[62] Accordingly, the Supreme Court was confronted with a litigation that highlighted the policy-related concerns that would arise from a decision to invest Rule 25 substitutions with potentially dispositive jurisdictional significance, namely, the loss of an otherwise binding judgment obtained through a long and costly judicial process. Finally, the *Freeport-McMoRan* Court also confronted a case in which the jurisdictional difficulties that arose from the Rule 25 substitution at issue were asserted only belatedly, tactically, and in an obvious attempt to forum shop.[63]

---

58. *Freeport-McMoRan*, 498 U.S. at 428-29.

59. *Id.* at 427.

60. McMoRan Oil & Gas Co. v. KN Energy, Inc., 907 F.2d 1022, 1023 (10th Cir. 1990).

61. *Id.* The quoted language reflects the then-prevailing jurisdictional doctrine. *See supra* note 34.

62. *McMoRan*, 907 F.2d at 1023 (noting that the trial court had entered a judgment in favor of McMoRan and FMPO for over $1.6 million).

63. *Id.* The procedural happenstance that confronted Freeport-McMoRan (and that ultimately required it to defend its judgment all the way to the Supreme Court) was intensified by the fact that at the time the district court denied the request to dismiss for lack of jurisdiction,

Freeport-McMoRan thus perfectly exemplified a case in which a dismissal for lack of federal subject matter jurisdiction based upon a post-litigation transfer of interest entirely irrelevant to the merits would either result in a remand to state court—thereby largely negating two years of substantial federal discovery and adjudication—or that would effectively deter parties from engaging in presumptively beneficial underlying economic transactions due to the fear of attendant jurisdictional consequences in the unrelated litigation. The Supreme Court's reluctance to embrace such a result, especially in a case in which a jurisdictional dismissal would negate the finality of a previously entered judgment on the merits, is not particularly surprising, especially given the facial doctrinal validity of the Court's holding.

C. *The Extension*

The practical and policy-centered consequences that would arise from the investiture of jurisdictional significance to a Rule 25 substitution are evident not only in *Freeport-McMoRan*, but also in the subsequent interpretation and extension of the Court's holding by the Court of Appeals. The leading case in this regard is the D.C. Circuit Court's decision in *Burka v. Aetna Life Insurance Co.* ("*Burka II*").[64] *Burka II* exemplifies the type of Rule 25(c) substitution addressed in *Freeport-McMoRan* and the breadth the United States Courts of Appeal typically accord to this holding.

The D.C. Circuit's opinion in *Burka II* arose from a foreclosure sale in which Aetna Life Insurance Company ("Aetna") acquired property that had previously been owned by a trust controlled by Paul and Robert Burka.[65] Shortly after purchasing the property from the Burka Trust, Aetna arranged to sell the property to American University ("American"), which intended to use the land to support its law school.[66] However, prior to the transfer to American, the Burkas brought an action against Aetna in D.C. Superior Court challenging Aetna's purported ownership of the property and the propriety of a commercial building located thereupon.[67] Aetna, a Connecticut corporation, promptly removed the lawsuit filed by the Burkas (citizens of Maryland and D.C.,

---

the Court had not yet held that partnerships were to be deemed citizens of each and every state of their constituent members, a holding that clearly arose only from a case rendered while *McMoRan* was pending before the Tenth Circuit. *See* discussion *supra* note 33. Freeport-McMoRan accordingly had little inkling that the marginal benefits of its proposed Rule 25 substitution would create massive procedural wrangling, including a defeat before the Tenth Circuit and the need for further legal expenditures in the Supreme Court.

64. Burka v. Aetna Life Ins. Co. (*Burka II*), 87 F.3d 478 (D.C. Cir. 1996).

65. *Id.* at 479.

66. *Id.* The transfer of the property from Aetna to American was made, as in *McMoRan*, for reasons entirely unrelated to the litigation. *Id.* at 481.

67. *Id.* at 479.

respectively) to federal court based upon the presence of diversity jurisdiction.[68] Aetna subsequently obtained summary judgment in the federal action.[69]

American had entered into a formal agreement with Aetna to purchase the underlying property before the Burkas filed their Superior Court action and Aetna's subsequent removal.[70] The Burkas were nonetheless content to remain in the district court to litigate against only Aetna, the existing record owner of the property.[71] Moreover, Aetna did not actually transfer the property to American until after the district court entered judgment in its favor.[72] In the meantime, the Burkas appealed the district court's adverse decision to the D.C. Circuit, which affirmed the judgment in virtually all respects, except for a minor claim that had been overlooked by the district court.[73]

Only upon remand from the D.C. Circuit did the case—for jurisprudential purposes—become interesting. First, prior to an evidentiary hearing on the remaining claim, in light of the post-filing transfer of the property from Aetna to American—as well as the fact that Aetna had almost entirely prevailed in both the district court and the D.C. Circuit[74]—Aetna moved under Rule 25(c) to add American as an additional defendant.[75] Because of the transfer of the property at issue, Aetna's strategy may have been to ensure that American would obtain the benefit of claim preclusion from Aetna's successful litigation.[76] Second, one week after Aetna filed a motion to substitute, the

---

68. *Id.*

69. *Id.*

70. *Id.*

71. *See id.*

72. *Id.*

73. *See* Burka v. Aetna Life Ins. Co. (*Burka I*), 56 F.3d 1509 (D.C. Cir. 1995) (per curiam).

74. It is unclear why Aetna thought that such a substitution was either necessary or beneficial, as American would have had the benefit of nonmutual defensive issue preclusion even absent joinder as a party. *See* Parklane Hosiery Co. v. Shore, 439 U.S. 322, 327-28 (1979). Although Aetna's tactics ultimately did not invalidate its successful litigation, they nonetheless resulted in substantial unnecessary expense and delay, including the need to defend multiple procedural appeals. *See infra* text accompanying notes 79-80.

75. *Burka II*, 87 F.3d at 479.

76. *See id.* Aetna did not seek to formally substitute American as the sole defendant as authorized by Rule 25, in part because Aetna held a mortgage on the underlying property and hence remained an interested party, but rather sought to join American as an additional defendant. *See id.* at 479 n.1. Rule 25 expressly permits both "pure" substitution as well as the joinder of the additional party as requested by Aetna. *See* FED. R. CIV. P. 25(c) (authorizing district courts to "direct[] the person to whom the interest is transferred to be substituted in the action or joined with the original party").

Aetna's decision to request that American be "joined" in the action rather than substituted was perhaps another tactical error, and it likely elicited the Burkas' subsequently filed motions as well as substantially strengthened the Burkas' claim that the "joinder" of an additional defendant divested the district court of subject matter jurisdiction notwithstanding the Supreme Court's—arguably distinguishable—holding in *Freeport-McMoRan*. *See supra* text

Burkas recognized a potential way out of their disastrous federal litigation against Aetna.[77] They sought to add American as a defendant pursuant to Rule 19 and remand the case to D.C. Superior Court pursuant to 28 U.S.C. § 1447(e) because the joinder of American (a citizen of the District of Columbia) would destroy diversity.[78]

The district court granted Aetna's motion to substitute American pursuant to Rule 25(c), denied the Burkas' motion to join American under Rule 19, and refused to remand for lack of subject matter jurisdiction, notwithstanding the undisputed lack of complete diversity amongst the current parties.[79] Judge Richey nonetheless recognized the novelty of the resulting jurisdictional challenge and, pursuant to a request by the Burkas, certified the matter for a possible interlocutory appeal.[80]

The D.C. Circuit granted the petition for review,[81] ordered a rapid schedule of briefing and oral argument,[82] and affirmed the district court's decision.[83]

---

accompanying notes 36-41. American's apparent desire to file a counterclaim against the Burkas based upon their alleged conduct during the pendency of the initial appeal hardly seems to have been worth the substantial delay and expense created by the motion to substitute. *See Burka II*, 87 F.3d at 480; *cf.* discussion *infra* Part III.B-C (discussing the procedural abuses that may improperly arise when substituted parties seek to assert additional claims against existing litigants notwithstanding the absence of original federal subject matter jurisdiction).

77. *See Burka II*, 87 F.3d at 479-80.

78. *See id.*; *see also* 28 U.S.C. § 1447(c) (2000) ("If at any time before final judgment it appears that the district court lacks subject matter jurisdiction, the case shall be remanded."); *id.* § 1447(e) ("If after removal the plaintiff seeks to join additional defendants whose joinder would destroy subject matter jurisdiction, the court may deny joinder, or permit joinder and remand the action to the State court."); *cf. infra* Part III.A.2 (discussing the holding of *Freeport-McMoRan* as applied to non-Rule 25 joinder in both removal and non-removal cases).

79. *See* Burka v. Aetna Life Ins. Co., 894 F. Supp. 28, 30-32 (D.D.C. 1995) (The court found that American was now already a defendant pursuant to Rule 25 and hence need not be further "joined."); *see also Burka II*, 87 F.3d at 480 (describing these holdings). The district court also simultaneously granted the motion of American to amend and supplement the answer in order to assert new affirmative claims against the Burkas, thereby authorizing the—somewhat unprecedented—federal adjudication of state law litigation between nondiverse parties. *See Burka II*, 87 F.3d at 480; *see also* discussion *infra* Part III.C (critically exploring the jurisdictional validity of such adjudication).

80. *See Burka*, 894 F. Supp. at 33.

81. *See Burka II*, 87 F.3d at 481-82.

82. The D.C. Circuit granted the petition for appeal on December 1, 1995, heard oral argument on the matter on May 9, 1996, and issued an opinion on June 14, 1996. *See id.* at 478, 482; *cf. supra* note 37 (noting a similarly rapid adjudication of *Freeport-McMoRan* by the Supreme Court). The prompt adjudications of *Burka II* and *Freeport-McMoRan* tend to support a conclusion that federal appellate judges generally—and perhaps not surprisingly—do not wish to linger on relatively arcane matters of civil procedure or feel particularly compelled to do so by the substantive intricacy of the underlying issues, at least when those matters involve Rule 25.

83. *See Burka II*, 87 F.3d at 482-84 ("In short, we find that the judgment of the District Court is exactly right.").

Chief Judge Edwards' opinion in *Burka II* not only concluded that substitution pursuant to Rule 25(c) was proper in the instant case,[84] but also expanded the core jurisdictional holding of *Freeport-McMoRan* in several different respects.

First, the D.C. Circuit held for the first time that the substantive dictates of Rule 25[85] and the jurisdictional rule articulated by the Supreme Court in *Freeport-McMoRan*[86] applied to cases originally filed by the plaintiff in federal court (e.g., the one filed by Freeport-McMoRan),[87] and to actions removed from state court.[88] Chief Judge Edwards' opinion thus (rightly) removed an arguable basis on which future litigants might attempt to distinguish *Freeport-McMoRan*,[89] notwithstanding the special statutory provisions that regulate removal and that might be viewed as inconsistent with some forms of substitution under Rule 25.[90]

---

84. *See id.* at 482 (characterizing the Burkas' contention that Rule 25(c) was facially inapplicable as "patently meritless").

85. *See id.* ("[T]he facts of this case fall squarely within the language and purpose of Rule 25(c), and we find no basis for any suggestion that Rule 25(c)'s applicability should be avoided merely because this is a removal case.").

86. *See id.* (finding that "there is no basis for appellants' contention that the holding of *Freeport-McMoRan* does not extend to removal cases" and arguing that "under *Freeport-McMoRan*, it is clearly understood that, as a general rule, the addition of a non-diverse party pursuant to Rule 25(c) does not deprive the District Court of subject matter jurisdiction," and that "[t]his general rule is equally applicable both in removal cases and cases originally brought in federal court").

87. *See* Freeport-McMoRan Inc. v. K N Energy, Inc., 498 U.S. 426, 427 (1991) (per curiam).

88. *See Burka II*, 87 F.3d at 482; *see also* Harry M. Gruber, *Civil Procedure: Adding a Nondiverse Party Pursuant to Federal Rule of Civil Procedure 25(c)*, 65 GEO. WASH. L. REV. 633, 636 (1997) (noting that in *Burka II* the D.C. Circuit "f[ound] no reason to differentiate between diversity cases removed to federal courts and those originating in federal court").

89. *See* Harry M. Gruber, *supra* note 88, at 633 (noting that *Burka II* "clarifies many issues surrounding the addition of a nondiverse party to a suit" as well as the interaction between Rule 25 and the statutory authorities that govern removal); *id.* at 637 ("The D.C. Circuit's opinion clarifies many of the issues surrounding the application of Rule 25(c), providing a more substantial and meaningful interpretation of . . . Rule 25(c) . . . in the light of *Freeport-McMoRan*."); *id.* ("This decision marks an advancement in interpreting the rules and statutes dealing with the addition of a party to a suit."). Mr. Gruber's discussion of *Burka II* does little more than re-publish the facts and holding of the case, and it does not critically examine either the validity or breadth of the D.C. Circuit's decision. *See also Law and Motion: Removal – Diversity Jurisdiction – Adding Nondiverse Parties*, FED. LITIGATOR, Nov. 1996 at 247, 248 ("*Burka* makes clear that the *Freeport-McMoRan* rule applies to removed cases, as well as cases originally brought in federal court.") (lacking critical exploration).

90. *See infra* text accompanying notes 91-104 (discussing the interaction between Rule 25 and 28 U.S.C. § 1447(e)).

Second, and more controversially, *Burka II* held that the provisions of 28 U.S.C. § 1447 were largely inapplicable to parties joined pursuant to Rule 25.[91] Section 1447(e) provides that "[i]f after removal the plaintiff seeks to join additional defendants whose joinder would destroy subject matter jurisdiction, the court may deny joinder, or permit joinder and remand the action to the State court,"[92] thereby generally precluding the effective post-removal joinder of an additional nondiverse defendant in federal diversity cases.[93] When enacting § 1447(e), Congress assuredly contemplated that the "joinder" referred to therein would be accomplished through Rule 19 ("Joinder of Persons Needed for Just Adjudication") or Rule 20 ("Permissive Joinder of Parties").[94] The addition of American pursuant to Rule 25(c) in *Burka II*, although formally referred to as a form of "substitution,"[95] nonetheless seemed to constitute "joinder" regulated by § 1447(e), if only due to the express use of the term "joinder" in Rule 25(c) to describe the addition of such a party.[96] In *Burka II*, the fact that the addition

---

91. *Burka II*, 87 F.3d at 483 ("Like the District Court, we are not persuaded by appellants' unsupported contention that a Rule 25(c) addition itself triggers the application of section 1447(e)'s remand requirement in this case.").

92. 28 U.S.C. § 1447(e) (2000).

93. *See* Mayes v. Rapoport, 198 F.3d 457, 461-62 (4th Cir. 1999) ("When a plaintiff seeks to join a nondiverse defendant after the case has been removed, the district court's analysis begins with 28 U.S.C. § 1447(e), which provides the district court with two options: 'If after removal the plaintiff seeks to join additional defendants whose joinder would destroy subject matter jurisdiction, the court may deny joinder, or permit joinder and remand the action to the State court.' These are the only two options for a district court faced with a post-removal attempt to join a nondiverse defendant; the statute does not allow a district court to retain jurisdiction once it permits a nondiverse defendant to be joined in the case."); *see also* 28 U.S.C. § 1447(c) ("If at any time before final judgment it appears that the district court lacks subject matter jurisdiction, the case shall be remanded.").

94. The legislative history of 28 U.S.C. § 1447(e) is bereft of any reference to Rule 25. Given the current lack of academic or judicial attention to Rule 25, it seems unlikely that Congress had expressly intended to exclude joinder under Rule 25 and apply § 1447 only to joinder pursuant to Rules 19 and 20 (and perhaps Rule 22). Of course, § 1447 is not necessarily inapplicable to Rule 25, but nonetheless helps to explain the relative complexity of the holding on this issue in *Burka II* as well as the precedential and practical significance of that determination.

95. *See* FED. R. CIV. P. 25 (entitled "Substitution of Parties"); *see also* Freeport-McMoRan Inc. v. K N Energy, Inc., 498 U.S. 426, 427 (1991) (per curiam) (describing the addition of FMPO as a proposed "substitut[ion]" even though FMPO was simply included as an additional plaintiff and Freeport-McMoRan continued to be a party).

96. *See* FED. R. CIV. P. 25(c) (expressly stating that a party added pursuant to Rule 25(c) may either "be substituted in the action or *joined* with the additional party") (emphasis added); *see also Burka II*, 87 F.3d at 479 n.2 (implicitly recognizing, albeit in an backhanded manner, that the addition of American might qualify as a "joinder" as that term is employed in § 1447(e)); *but see id.* at 483-84 (stating that "we . . . find nothing in the law to suggest that a transfer-of-interest-based substitution or addition under Rule 25(c) is a form of 'joinder' within the meaning of section 1447(e)"); Gruber, *supra* note 88, at 636 (stating that "[t]he D.C. Circuit decided not to remand in [*Burka*] . . . [because] the court explained that Rule 25(c) does not

of American was ultimately accomplished at the request of an existing defendant[97]—thereby making facially inapplicable the provisions of 28 U.S.C. § 1447(e), which expressly apply only when "the *plaintiff* seeks to join additional defendants whose joinder would destroy subject matter jurisdiction"[98]—would be a slender basis to conclude that § 1447 does not require a remand,[99] particularly given the fact that 28 U.S.C. § 1447(c) compels an identical remand to state court regardless of which party causes the resulting

---

involve 'joinder' of a party, triggering section 1447(e)'s remand requirement, but only involves the 'addition' of a party, which does not trigger section 1447(e)").

The potential comment from *Burka II* that allegedly distinguished "joinder" governed by § 1447(e) as opposed to a purportedly distinct "addition" under Rule 25 was unsupported by any authority or argument apart from mere citation to *Freeport-McMoRan*. Yet, *Freeport-McMoRan* did not involve a removed action nor did it attempt in any manner to interpret the textual dictates of § 1447(e); accordingly, it wholly fails to support the D.C. Circuit's conclusion in this regard. Moreover, the *Burka II* court did not explain how a purported distinction between "addition" and "joinder" would comport with the deliberate decision in Rule 25(c) to employ the latter term instead of the former. Finally, any purported distinction between joinder and substitution would be the worst of all possible bases upon which to found the judgment in *Burka II*.

The fact that a particular type of Rule 25(c) substitution may indeed qualify as a regulated "joinder" governed by § 1447(e) does not mean that all such substitutions equally qualify, a fact previously overlooked. Rule 25(c) is the only component of Rule 25 that facially authorizes the "join[der]" of an additional party. By contrast, Rules 25(a), 25(b), and 25(d) each expressly refer only to the "substitution" of a party. Although the proper result is by no means certain, perhaps § 1447(e) applies only to parties "joined" (not substituted) pursuant to Rule 25(c) or even that § 1447(e) is entirely inapplicable to Rules 25(a), (b) and (d). *See also infra* note 100 (discussing additional reasons why § 1447(e) might not apply the addition of a particular party pursuant to Rule 25(c)); *cf. infra* note 100 (exploring nonetheless the potential applicability of 28 U.S.C. § 1447(c) as requiring a remand even if § 1447(e) does not apply).

97. *See Burka II*, 87 F.3d at 479-80; *cf.* CHARLES ALAN WRIGHT, ARTHUR R. MILLER, & MARY KAY KANE, *supra* note 19, § 1958 ("The motion for substitution [under Rule 25] may be made by any party."); *see also* Montecatini Societa Generale per L'Industria Mineraria e Chimica v. Humble Oil & Ref. Co., 261 F. Supp. 587, 591 (D. Md. 1966) (noting identical rule).

98. 28 U.S.C. § 1447(e) (emphasis added). This fact did not go unnoticed by the D.C. Circuit in *Burka II*, which mentioned this allegedly distinguishing characteristic in holding that § 1447 did not require the district court to remand the action upon the joinder of American as a defendant. *See Burka II*, 87 F.3d at 483.

99. Such a limited holding might, *inter alia*, suggest that a remand is required when the plaintiff moves for substitution but not when the defendant so moves, a distinction that would appear to meaninglessly elevate form over substance because it is the underlying factual predicates—the merger or assignment—that typically compel the substitution. Such a holding would also create a seemingly untenable—and perhaps unfair—doctrine in which the parties had an unequal ability to obtain the benefits of Rule 25. *See also* discussion *infra* Parts III.B-C (discussing similar unsupportable doctrinal canons associated with the regime articulated by *Freeport-McMoRan* that arise from the differential treatment of Rules 19, 24, and 25).

lack of subject matter jurisdiction.[100] Accordingly, the D.C. Circuit in *Burka II* squarely held[101] that the remand potentially required by § 1447 was

---

100. *See* 28 U.S.C. § 1447(c) ("If at any time before final judgment it appears that the district court lacks subject matter jurisdiction, the case shall be remanded."); *see also* FED. R. CIV. P. 12(h)(3) ("Whenever it appears by suggestion of the parties or otherwise that the court lacks jurisdiction of the subject matter, the court shall dismiss the action."). The only substantive difference between §§ 1447(c) and 1447(e), as implicated by the request for a remand in *Burka II*, is that the latter expressly gives the court discretion to refuse the requested joinder and hence retain jurisdiction, while the former compels a mandatory remand. This is a distinction without a difference, as litigants can just as easily obtain a remand under § 1447(c) as they can under § 1447(e). Moreover, as applied to cases arising under Rule 25, the difference between §§ 1447(c) and 1447(e) is wholly irrelevant, as Rule 25 already imparts virtually unlimited discretion on the district court to either grant or deny the requested substitution, thereby incorporating—albeit in a rule-based as opposed to statutory context—the precise type of allegedly "distinguishing" discretion present in § 1447(e) but formally absent in § 1447(c). *See, e.g.*, Finova Capital Corp. v. Lawrence, No. 399CV2552-M, 2000 WL 1808276, at *1 (N.D. Tex. Dec. 8, 2000) ("[S]ubstitution is never mandatory. Rather, the district court has broad discretion to order substitution, deny substitution, or direct that the transferee be joined as an additional party.") (citations omitted); Jazzland, Ltd. v. Bordenave, No. CIV. A. 98-1356, 1999 WL 243820, at *1 (E.D. La. Apr. 22, 1999) (noting similar rule); Barker v. Jackson Nat'l Life Ins. Co., 163 F.R.D. 364, 366 (N.D. Fla. 1995) (same); *see also* Adams & Arguin, *supra* note 20, § 14.7(a)(2) ("The disposition of a motion pursuant to FRCP [sic] 25 is committed to the sound discretion of the trial court. Substitution of parties under the Rule is never mandatory; it is 'purely a matter of convenience.'") (citation omitted); CHARLES ALAN WRIGHT, ARTHUR R. MILLER, & MARY KAY KANE, *supra* note 19, § 1958 ("An order of joinder is merely a discretionary determination by the trial court that the transferee's presence would facilitate the conduct of the litigation.").

The conflict between § 1447 and Rule 25 cannot properly be resolved, at least in most cases, merely by reference to the apparent textual limitations of 28 U.S.C. § 1447(e), especially in light of the contemporary provisions of both 28 U.S.C. § 1447(c) and Rule 12(h)(3). Regardless of the particular statutory language employed, if Rule 25 joinder properly divests the district court of jurisdiction, remand is required; if not, the case may remain in federal court. The jurisdictional significance of Rule 25 joinder is dispositive, not the particular (plaintiff-focused) language of § 1447(e).

101. Chief Judge Edwards' opinion in *Burka II* somewhat addresses this issue. While also viewing § 1447(e) inapplicable to joinder pursuant to Rule 25, the district court alternately ruled that it would exercise its substantial discretion pursuant to § 1447(e) not to allow joinder were such an event to require a remand. *See Burka II*, 87 F.3d at 481 n.5; *see also* Wyant v. Nat'l R.R. Passenger Corp., 881 F. Supp. 919, 923 (S.D.N.Y. 1995) ("Under Section 1447(e) courts have discretion either to deny joinder of a non-diverse defendant and retain jurisdiction, or to permit joinder and remand the case to state court."). Thus there was no doubt that the Burkas would lose their motion to remand the case to state court, and the only matter that was really at stake was American's largely irrelevant ability to be a formal party to the action. As a result, given the virtually certain persistence of the prior judgment in *Burka I*, as well as the continuing ability of American to bind the Burkas to whatever judgment ultimately resulted from *Burka II*, the D.C. Circuit's decision to grant the petition for interlocutory appeal and to rely on the categorical inapplicability of § 1447 to Rule 25, rather than the district court's alternate

categorically inapplicable to parties joined pursuant to Rule 25(c),[102] regardless of whether the joinder of these same parties under a different Rule—Rules 19 or 20—would require a remand to state court.[103] The court's conclusion not only extended the practical and doctrinal import of the Supreme Court's original holding in *Freeport-McMoRan*, but also resolved a complicated, and previously unaddressed, procedural concern that could otherwise have served to substantially limit the ability to effectively employ Rule 25 substitution in both removed and other types of cases.[104]

Finally, the D.C. Circuit's momentous opinion addressed for the first time a situation in which a party added pursuant to Rule 25 sought to interject itself into the existing dispute and to assert an additional claim against an adverse party,[105] a claim over which the district court had only supplemental jurisdiction.[106] Yet, the D.C. Circuit did not view such an unprecedented event as having any substantial significance, and held without any substantive analysis[107] that supplemental jurisdiction existed over such ancillary[108] claims

---

discretionary decision not to remand, tends to suggests the court's desire to adjudicate the core jurisdictional issue.

102. *See Burka II*, 87 F.3d at 484 (agreeing with the district court and holding that, pursuant to the Supreme Court's decision in *Freeport-McMoRan*, "'[t]he option of remanding the case [provided in section 1447(e)] is . . . not available to the Court if A[merican] U[niversity] is substituted in as a party defendant pursuant to Rule 25(c).'" (quoting *Burka I*, 894 F. Supp. at 29)) (first and second alteration in original).

103. *See Burka II*, 87 F.3d at 484 ("[W]e find nothing in the law suggesting that either Rule 19 or section 1447(e) trumps Rule 25(c) when all may be applicable, especially when the Rule 25 motion is filed first in time. The District Court certainly was within its discretion in considering the Rule 25(c) issue first, and, having properly found Rule 25(c) applicable, the court was in no way obligated to apply some other method of party addition that would necessitate a remand."); *see also id.* at 479 n.2 (arguing that "[t]he application of each rule [to 28 U.S.C. § 1447] is . . . distinct"); *infra* Part III.A.2 (critically examining such differential treatment).

104. *Cf.* discussion *infra* Part III.C (discussing various attempts to substitute that would be foreclosed by the D.C. Circuit's analytical framework).

105. *See Burka II*, 87 F.3d at 479-80 (noting American's assertion, upon being added as a party defendant, of affirmative claims against the Burkas). By contrast, when FMPO was joined in *Freeport-McMoRan*, FMPO sought only to continue to assert the same claims originally brought by the initial plaintiff and did not raise additional claims of its own. *See* Freeport-McMoRan Inc. v. K N Energy, Inc., 498 U.S. 426, 427 (1991) (per curiam); *see also* discussion *infra* Part III.C (discussing the significance of and problems created by the inclusion of additional claims by the substituted party).

106. *See* 28 U.S.C. § 1367 (2000).

107. *See Burka II*, 87 F.3d at 480 n.4 (resolving this issue in one unsupported sentence contained in a single footnote of the opinion).

108. The term "ancillary," is not intended to incorporate the common law legal definition of this term, but rather to indicate that the additional claims raised by American upon entry as a plaintiff pursuant to Rule 25(c) were merely post-filing assertions related to the original complaint.

and that 28 U.S.C. § 1367(b) did not deprive the district court of jurisdiction.[109] The court's conclusion again extended both the doctrinal and practical breadth of the Supreme Court's opinion in *Freeport-McMoRan*, and also substantially advanced the power of party substitutions accomplished pursuant to Rule 25.[110] The opinion of the D.C. Circuit accomplished additional functions as well, each of which expanded the breadth and significance of *Freeport-McMoRan*.[111]

Other less critical district and appellate court applications of Rule 25 have also expanded *Freeport-McMoRan*.[112] As a result, the central jurisdictional holding in *Freeport-McMoRan* presently authorizes limited, "pure" substitution of one party for another in cases originally filed in federal court and extends both the form—joinder and removal—and substantive content—inclusion of additional supplemental claims—of substitution under Rule 25. As currently interpreted and applied, Rule 25 does more than merely authorize the formal replacement of Gail Norton for Bruce Babbit as the formal party defendant in suits against the Secretary of the Department of Interior[113] or substitution of Mrs. Reagan for former President Reagan upon the death or legal incompetence of the latter.[114] The "easy" holding in the Supreme Court's unanimous per curiam opinion in *Freeport-McMoRan* is similarly not as facile.

D. The Caveat

Before addressing the central difficulties associated with the contemporary application of Rule 25 and the Supreme Court's holding in *Freeport-McMoRan,* one must recognize the final manner in which the Court may have slightly limited—generally correctly[115]—the scope of Rule 25 as applied to the post-filing substitution of nondiverse parties. While the Supreme Court ratified the particular substitution at issue in *Freeport-McMoRan,* it simultaneously appeared to articulate—albeit in an obtuse and backhanded manner—a limiting principle on the scope of such permissible conduct. The Court stated:

---

109. *See Burka II*, 87 F.3d at 480 n.4; *see also* discussion *infra* Part III.C (critically examining both the wisdom and propriety of this core jurisdictional holding).

110. *See* discussion *infra* Part III.C.

111. *See, e.g., infra* note 119 (discussing the somewhat limited application in *Burka II* of the "indispensable party" caveat of *Freeport-McMoRan*); *infra* note 213 (exploring the *Burka II* court's differential treatment of joinder pursuant to Rule 25(c) as opposed to alternative forms of joinder).

112. *See infra* notes 122-23 and accompanying text.

113. *See* FED. R. CIV. P. 25(d) (authorizing substitution of official defendants); James v. Norton, 176 F. Supp.2d 385, 387 n.1 (E.D. Pa. 2001) (granting such a substitution).

114. *See* FED. R. CIV. P. 25(a), 25(b) (authorizing, respectively, substitution upon death and incompetence).

115. *But see infra* text accompanying notes 126-28 (noting the lack of conceptual foundation for this rule); discussion *infra* Part III.D (discussing both over- and under-inclusive components of the Court's holding).

> The opinions of the District Court and the Court of Appeals establish that the plaintiffs and defendant were diverse at the time the breach-of-contract action arose and at the time that federal proceedings commenced. The opinions also confirm that FMPO was not an "indispensable" party at the time the complaint was filed; in fact, it had no interest whatsoever in the outcome of the litigation until sometime after suit was commenced. Our cases require no more than this.[116]

The Court's reference in *Freeport-McMoRan* to the essential status of FMPO as a "dispensable" party is classic dicta, and concerned an issue not raised by the parties, not implicated by the facts of the case, and not addressed by any of the courts below.[117] Accordingly, several initial commentators were somewhat hesitant in their recognition of the Court's "apparent" limitation in on the core jurisdictional holding in *Freeport-McMoRan*.[118] The exception seemed to render the general rule of *Freeport-McMoRan* inapplicable to the proposed substitution of parties initially indispensable[119] at the time the action was filed. Preliminary academic uncertainty aside, contemporary courts and commentators now uniformly view the Court's opinion in *Freeport-McMoRan* to have squarely articulated, in categorical language,[120] a rule that federal subject matter jurisdiction exists over nondiverse parties added pursuant to Rule 25 only when the new party to be substituted was not originally indispensable.[121] Post-*Freeport-McMoRan* federal courts have similarly and repeatedly recognized[122] and expressly applied[123] this "indispensable party" limitation, and accordingly its present doctrinal persistence is certain.

---

116. Freeport-McMoRan Inc. v. K N Energy, Inc., 498 U.S. 426, 428 (1991) (per curiam).

117. *See id.* (inserting this statement in the midst of an otherwise standalone paragraph); *see also* McMoRan Oil & Gas Co. v. K.N. Energy, Inc., 907 F.2d 1022, 1022-25 (1990) (failing to contain any discussion of dispensable or indispensable parties).

118. *See, e.g.*, Gruber, *supra* note 88, at 636 (indicating that the Supreme Court in *Freeport-McMoRan* had "alluded" to a possible limitation for initially indispensable parties).

119. It is somewhat unusual that the Supreme Court even chose to employ the concept of "indispensable" parties. Rule 19 expressly prefers the somewhat more descriptive nomenclature of "Persons to Be Joined if Feasible" and "Determination[s] by [the] Court Whenever Joinder [is] Not Feasible". FED. R. CIV. P. 19(a)-(b). Nevertheless, there is a continuing presence in Rule 19(b) of the term "indispensable". *See* FED. R. CIV. P. 19(b); *see also* discussion *infra* Part III.D (discussing additional substantive problems that may arise, at least in part from the Court's use of the common law appellation).

120. *See McMoRan*, 498 U.S. at 428 (stating that "[o]ur cases require no more" than the new party to be joined pursuant to Rule 25 was not initially indispensable).

121. *See, e.g.*, Adams & Arguin, *supra* note 20, § 14.7(a)(3); DAVID HITTNER ET AL., FEDERAL CIVIL PROCEDURE BEFORE TRIAL ¶ 2:379 (5th Cir. ed. 2006); JAMES WM. MOORE ET AL., MOORE'S FEDERAL PRACTICE ¶ 25.33 (3d ed. 2006).

122. *See, e.g.*, Whalen v. Carter, 954 F.2d 1087, 1096 (5th Cir. 1992) (noting "the Court in *Free[p]ort*[-]*McMoRan* concluded that the addition of a nondiverse party does not defeat diversity jurisdiction unless the party was indispensable at the time the plaintiff filed his complaint"); *see also* Cobb v. Delta Exports, Inc., 186 F.3d 675, 679 (5th Cir. 1999); Sharp v. KMart Corp., 991 F. Supp. 519, 523 (M.D. La. 1998).

Yet, the foundation upon which the Court in *Freeport-McMoRan* articulated this jurisdictional restriction remains almost entirely unclear.[124] The Supreme Court established this limitation merely by stating, *ex cathedra*, that FMPO was not an originally indispensable party and that "[o]ur cases require no more than this."[125] It is difficult to discern the jurisprudential basis upon which this statement relies. The Supreme Court entirely failed to cite or even refer to any of these alleged "cases,"[126] and no such authority in fact exists.[127] The doctrinal foundation for the *Freeport-McMoRan* Court's "indispensable party" limitation accordingly remains fairly indeterminate.[128] Nonetheless, the

---

123. *See, e.g.*, *Burka II*, 87 F.3d 478, 482-83 (D.C. Cir. 1996) (applying squarely the "indispensable party" limitation of *Freeport-McMoRan* in a Rule 25(c) case and affirming substitution on the grounds that the party to be joined was not indispensable at the time the lawsuit was initially filed); Kerr v. Smith Petroleum, 889 F. Supp. 892, 894-96 (E.D. La. 1995) (recognizing and applying this same doctrinal limitation and affirming joinder of party on the grounds that it was not initially indispensable); *see also* Butcher v. Hildreth, 992 F. Supp. 1420, 1423 (D. Utah 1998) (distinguishing *Freeport-McMoRan* and remanding the action to state court because, *inter alia*, the party to be joined was indispensable).

124. *Cf.* discussion *infra* Parts II.B-C (discussing at length the differential scope and breadth of this holding, depending on the particular doctrinal basis on which the Court intended to rely).

125. Freeport-McMoRan Inc. v. K N Energy, Inc., 498 U.S. 426, 428 (1991) (per curiam).

126. *See id.* (including only a *single* citation in the entire opinion subsequent to this statement—Owen Equipment & Erection Co. v. Kroger, 437 U.S. 365 (1978)—and, even then, expressly doing so only to argue that *Owen*, a case relied upon by the respondent, was distinguishable from the principle on which the Court relied); *but see infra* text accompanying notes 327-31 (supporting the Court's holding by articulating an analogy to *Owen* as applied to particular types of substitutions).

127. *See Freeport-McMoRan*, 498 U.S. at 428 (referring, *sub silentio*, to "our" cases); *see* discussion *infra* Part III.D (discussing the numerous jurisprudential difficulties attendant to the judicial and academic assumptions regarding the doctrinal basis for this limitation); *infra* text accompanying notes 148-58 (articulating an arguably—albeit only partially—valid foundation for such a restriction).

128. The Supreme Court in *Freeport-McMoRan* may perhaps have articulated this limitation based upon an entirely unarticulated rough analogy to the longstanding common law rule that the intervention of a nondiverse party would not defeat preexisting diversity jurisdiction unless the proposed intervenor was indispensable. *See* CHARLES ALAN WRIGHT, ARTHUR R. MILLER AND MARY KAY KANE, *supra* note 19, § 1917 (noting this common law rule). The Supreme Court may simply have derived the apparent "indispensable party" limitation from a mental comparison similar to the following: Rule 24 (Intervention) and Rule 25 (Substitution) = Pretty Much the Same?

The central problem with the validity of such a proposition—wholly apart from the plethora of doctrinal flaws that would necessarily arise, as discussed in Part III.D—is that Rule 25 fundamentally differs than Rule 24. Also, the common law "indispensable party" principle was abrogated by Congress in 28 U.S.C. § 1367(b). *See* 28 U.S.C. § 1367(b) (2000) (rejecting supplemental jurisdiction over claims by plaintiff against parties added pursuant to Rule 24); Lumber Ins. Cos. v. Allen, 892 F. Supp. 31, 33 (D.N.H. 1993) (holding that § 1367(b) precluded the proposed intervention of a nondiverse party as defendant).

Court was substantively correct in the view that Rule 25 should not typically authorize the substitution of initially indispensable parties.[129] For example, a plaintiff who would be otherwise barred from litigating alongside a co-plaintiff—because the co-plaintiff resided in the same state as one of the defendants—should not be allowed to circumvent the longstanding requirement of complete diversity[130] as well as the requirement in Rule 19(b)[131] by simply litigating alone, promptly assigning all or part of her claim to the co-plaintiff, and immediately moving to join the nondiverse, and otherwise indispensable, co-plaintiff pursuant to Rule 25(c).[132]

Accordingly, to the extent that *Freeport-McMoRan* articulates a common law limitation on the jurisdictional implications of substitution under Rule 25,[133] the inapplicability of *Freeport-McMoRan* to an indispensable co-plaintiff would be entirely justified,[134] and would also follow analogously from prior

---

The completely unsupported articulation of the indispensable party limitation yet again demonstrates why *Freeport-McMoRan* is unlikely to be viewed as the Court's intellectual zenith.

129. *See also infra* text accompanying notes 317-24 (articulating particular circumstances in which the Court's facially categorical principle does not properly apply).

130. *See* Strawbridge v. Curtiss, 7 U.S. 267, 267 (1806) (refusing to support diversity jurisdiction when the parties were from the same state).

131. *See* FED. R. CIV. P. 19(b) (requiring dismissal when the balance of the factors articulated therein weighs in favor of such a course of action).

132. *See* 28 U.S.C. § 1367(b) (2000) (precluding such a result as a matter of law where the co-plaintiff is to be joined pursuant to Rules 19 or 24); *cf. infra* text accompanying notes 236-51 (discussing related procedural abuses that might still arise even under the Supreme Court's existing limitations and accordingly calling for even more expansive limitations on joinder than presently exist). The abusive exemplars discussed above may, in fact, already be partially allowed when the plaintiff creatively employs Rule 20(a) rather than Rules 19 or 24. *See, e.g.*, Sunpoint Sec., Inc. v. Porta, 192 F.R.D. 716, 719 (M.D. Fla. 2000) (allowing precisely such activities). The appropriate response, at least for those who conclude that such types of joinder should not be allowed under *any* of the Rules, would be to include Rule 20 alongside Rule 25 in the various proposals for modification or discretionary application of supplemental jurisdiction advanced above. Much of what is said with respect to Rule 25(c) might apply equally to Rule 20.

133. *See also* discussion *infra* Part IV (discussing both contemporary and proposed statutory, non-common law restrictions).

134. *See* Acton Co. v. Bachman Foods, Inc., 668 F.2d 76, 79-80 (1st Cir. 1982) ("Applying these policies to the instant case, we conclude that Acton cannot be joined to this diversity action without destroying the court's subject matter jurisdiction. First, to allow Acton as co-plaintiff to bring its action against a non-diverse defendant would be to permit ACIM, Acton's wholly owned subsidiary, to create diversity jurisdiction simply by omitting Acton from the original complaint and then waiting for Acton to be joined under Rule 19. Second, given the relationship [as parent and subsidiary] between Acton and ACIM and the fact that they have substantially identical interests in the lawsuit, we are unable to view Acton as a party hailed into federal court against its will and thus entitled to invoke ancillary jurisdiction; rather, ACIM has chosen a federal forum, and neither ACIM nor Acton can equitably complain about the court's jurisdictional limitations."); *see also infra* notes 230-32 and accompanying text (articulating additional policy-related concerns in such settings). *But cf. infra* notes 233-40 and

common law jurisdictional limitations.[135] The preclusion of substitution pursuant to Rule 25 in such a setting would seem an appropriate restriction on the Court's otherwise expansive holding.[136] The common law "indispensable party" limitation would properly apply to the addition of a nondiverse co-plaintiff, and also to the post-filing substitution of an indispensable defendant. A contrary holding would allow circumvention of the complete diversity rule and the dictates of Rule 19(b).[137]

Take, for example, a situation in which a plaintiff from Maine is injured by the joint conduct of Monster, Inc. (also from Maine) and Tiny, Inc. (a citizen of Tennessee). The complete diversity rule would clearly preclude the plaintiff from initially filing a federal lawsuit that sought relief pursuant to state law and that named both Monster and Tiny as defendants.[138] Moreover, if Monster were an indispensable party,[139] the plaintiff could not secure the benefits of a federal forum even by seeking relief solely against Tiny, as Tiny could obtain a dismissal of any such action pursuant to Rule 19(b), notwithstanding the presence of complete diversity amongst the existing parties. Imagine, however, that Monster announces that it will soon acquire Tiny, subject to both

---

accompanying text (arguing that even *dispensable* co-plaintiffs should, in particular cases, also be barred from participation pursuant to Rule 25).

135. *See, e.g.*, Owen Equip. & Erection Co. v. Kroger, 437 U.S. 365, 376 (1978) (holding that the common law ancillary jurisdiction of federal courts did not extend to the adjudication of a claim by an original plaintiff against a nondiverse third-party defendant properly impleaded by the defendant pursuant to Rule 14); *Acton*, 668 F.2d at 79-80 (holding that federal common law jurisdiction did not extend to the post-filing joinder of a co-plaintiff that shared the citizenship of one of the existing defendants).

136. *Cf.* Hansen v. Peoples Bank of Bloomington, 594 F.2d 1149, 1153-54 (7th Cir. 1979) (dismissing action pursuant to Rule 19(b) because, although the plaintiff trust beneficiary was completely diverse from the defendant-trustee, other trust beneficiaries were both nondiverse and indispensable to the action).

137. *See supra* text accompanying notes 130-32.

138. *See* 28 U.S.C. § 1332(a) (2000).

139. Tortfeasor defendants are generally not indispensable parties as a mere result of their potential joint or several liability. *See, e.g.*, Temple v. Synthes Corp., 498 U.S. 5, 7 (1990); Stabilisierungsfonds Fur Wein v. Kaiser Stuhl Wire Distribs. Party Ltd., 647 F.2d 200, 207 (D.C. Cir. 1981). A more legally harmonious example of an indispensable defendant in a situation closely analogous to the one articulated above would be when a lessor plaintiff seeks to cancel a lease and names the diverse owner of the property, but not the nondiverse royalty holders, as a defendant. *See* Hilton v. Atl. Ref. Co., 327 F.2d 217, 219 (5th Cir. 1964) (dismissing such an action pursuant to Rule 19(b) on the grounds that the royalty holders were indispensable); *see also* Naartex Consulting Co. v. Watt, 722 F.2d 779, 785 (D.C. Cir. 1983); Vasser v. Shilling, 93 F.R.D. 146, 150 (D. La. 1982).

This Article employs the more commonplace example of defendant joint tortfeasors rather than the more esoteric, but legally more precise, situation of royalty owners for purposes of readability and to advance the critique discussed above. This Article argues that the *Freeport-McMoRan* "indispensable party" limitation is profoundly underinclusive, in part because this limitation would not typically preclude a Rule 25(c) substitution, and hence circumvention of the complete diversity rule, in the Monster and Tiny example discussed above.

shareholder and regulatory approval: plaintiff would, absent Rule 25(c) and *Freeport-McMoRan*, still be unable to obtain federal adjudication of her action. Plaintiff could not (1) file a pre-acquisition complaint against both Tiny and Monster in federal court, given the requirement of complete diversity, (2) file a pre-acquisition federal complaint against Tiny alone, given Rule 19(b); or (3) file either a pre- or post-acquisition federal complaint solely against Monster because subject matter jurisdiction over such state law claims would not exist.

In this context, the *Freeport-McMoRan* limitation becomes essential. The Court's general holding—absent the "indispensable party limitation"—would substantially alter preexisting jurisprudence. Allowing the plaintiff to file a federal lawsuit solely against Tiny, immediately prior to the acquisition, and to employ Rule 25(c) to substitute, or join,[140] the nondiverse Monster as a defendant shortly thereafter[141] would effectively circumvent each of the existing limitations on federal adjudication of such a lawsuit.[142] The inequitable,[143] doctrinally revolutionary,[144] and practically deleterious[145] consequences of such

140. *See supra* text accompanying note 132.

141. *Cf. infra* text accompanying notes 160-70 (discussing possible doctrinal arguments through which Monster might try, but—at least under existing jurisprudence—perhaps fail to defeat such a substitution absent the "indispensable party" limitation of *Freeport-McMoRan*). Plaintiff in this example would need to properly time her lawsuit so that it was simultaneously filed prior to the completion of the acquisition, hence authorizing substitution of Monster pursuant to a post-filing transfer of interest as required by Rule 25(c), while not long before this event. Otherwise, Tiny could successfully obtain a dismissal of the action pursuant to Rule 19(b) prior to the conclusion of the acquisition and the motion to substitute. Given that Rule 4(m) permits the plaintiff to file her lawsuit up to 120 days prior to serving Tiny, as well as the easy ability to file an action immediately prior to an acquisition, for example, filing a lawsuit a single day prior to the formal finalization of this event, the requisite temporal requirements would usually present little difficulty to a plaintiff in this situation. *See* FED. R. CIV. P. 4(m).

142. *See supra* notes 128-32 and accompanying text (discussing 28 U.S.C. § 1332(a), Rule 19, and the common law).

143. The situation discussed above would, absent the *Freeport-McMoRan* limitation, result in a practically dispositive "race to the courthouse" between plaintiff's filing a motion to substitute and defendant's filing a motion to dismiss for failure to join an indispensable party. It would also favor the type of tactically delayed—"sandbagged"—service of process and wealthier, repeat-player defendants who had the resources and incentive to obtain real-time information regarding the filing, as opposed to service, of a complaint.

144. *See supra* text accompanying notes 1-22 (discussing the contours of pre-*Freeport-McMoRan* and common law jurisprudence).

145. In situations in which such a combination would permit the plaintiff to obtain the benefits of a federal forum, Monster might, for example, tactically choose to delay the lucrative consummation of the Tiny acquisition until after Tiny filed a successful request for a Rule 19 dismissal, or until the underlying statute of limitations had expired. Moreover, in particularly economically close cases, the associated delay or possibility of deleterious, litigation-related consequences might result in a decision to entirely forego an otherwise beneficial transaction. *Cf.* Freeport-McMoRan Inc. v. K N Energy, Inc., 478 U.S. 426, 428-29 (1991) (per curiam) (articulating analogous policy-related consequences as support for the Court's general holding, though not facially recognizing the similar benefits of the equally important limitation at hand);

a result completely justify the central limitation, however backhanded and diffuse,[146] of the opinion in *Freeport-McMoRan* that refrains from extending the Court's jurisdictional[147] holding to the substitution of indispensable parties pursuant to Rule 25(c).

There is also an entirely legitimate basis for the "indispensable party" limitation in *Freeport-McMoRan*, wholly apart from the possible divestment of subject matter jurisdiction[148] in the event of Rule 25 substitution. Rule 25 is entirely procedural in nature,[149] expressly designed not to alter the substantive positions or defenses of the parties.[150] The effective Rule 25 substitution of an indispensable party, especially in cases similar to the exemplar previously discussed[151] would fundamentally alter the status of the litigation. Prior to substitution, the existing defendants would possess a compelling Rule 19(b) defense to the federal action. By contrast, after the substitution, the Rule 25 joinder of the previously indispensable party would presumably rescue an otherwise fatally defective action from dismissal.[152] Such a dispositive

---

Gruber, *supra* note 88, at 636 ("The [*Freeport-McMoRan*] Court reasoned that if anything less than the addition of an indispensable nondiverse party destroyed jurisdiction, then the lengthy course of litigation might inhibit normal business transactions concerning the property."); *infra* text accompanying notes 283-85 (discussing similar harmful policy-related consequences that continue to arise as a result of the underinclusive nature of the *Freeport-McMoRan* limitation).

146. *See supra* text accompanying note 116; *cf. infra* text accompanying notes 309-14 (attempting to impart doctrinal coherence and a definitive scope to this holding, and critically examining the contemporary interpretation of this limitation).

147. In *Freeport-McMoRan,* the Court appears to have founded this caveat on the particular jurisdictional difficulty at issue, namely the potential lack of post-substitution diversity. Hence, it did not facially extend the underlying limitation to situations in which substitution pursuant to Rule 25 would not create jurisdictional difficulties. *Cf. infra* text accompanying notes 303-44 (asserting both the underinclusive and overinclusive nature of this limitation and advancing a supplemental vision of this principle founded upon Rule-based rather than jurisdictional doctrines that may more effectively resolve the injurious consequences of the existing jurisprudential regime).

148. *See also infra* text accompanying notes 159-66 (discussing critically both the content and consequences of the potential jurisdictional basis of this limitation).

149. *See* Adams & Arguin, *supra* note 20, § 14.7(a)(1) ("The Rule is solely procedural and does not determine the substantive rights of parties in the situations to which it applies."); HITTNER ET AL., *supra* note 121, ¶ 7:356 ("This is purely a rule of procedure. The effect of death, incompetency, assignment, etc. is determined by applicable substantive law.").

150. *See, e.g.*, Covington Grain Co. v. Deal, 638 F.2d 1362, 1364 (5th Cir. 1981) ("Rule 25(c) is not designed to create new relationships among parties to a suit . . . ."); 59 AM. JUR. 2D *Parties* § 321 (2002) (noting that proper Rule 25 substitution "[is] procedural and do[es] not affect the substantive rights of the parties").

151. *See supra* text accompanying notes 138-47 (advancing the co-plaintiff, Monster and Tiny, and related examples in which Rule 25(c) might abusively be employed absent the Court's jurisdictional limitation on the substitution of indispensable parties).

152. After all, it would be difficult to argue that a party "could not be joined," thus requiring dismissal of the existing action pursuant to Rule 19(b), if after substitution, the dispositively absent party was in fact effectively joined. *Cf. infra* text accompanying notes 193-

alteration of the underlying litigation may well be inconsistent with the substantive dictates of Rule 25, which is neither "designed to create new relationships among parties to a suit"[153] nor permissibly utilized to change the existing substantive defenses of the parties.[154] These substantive limitations on Rule 25 could suggest that the Court's "indispensable party" limitation arises from the substantive structure and purpose of Rule 25 itself,[155] and not as a common law restriction on the assertion of federal diversity jurisdiction.[156]

In *Freeport-McMoRan,* the Supreme Court articulated a variety of valuable doctrinal, policy-related, and efficacious jurisprudential dictates. Notwithstanding the plethora of substantive and stylistic errors,[157] the Court properly resolved the case before it. Furthermore, it issued an opinion regarding the jurisdictional effect of Rule 25(c) substitution, the central features of which appear entirely reasonable.[158]

III. THE PROBLEMS OF THE EXISTING JURISPRUDENTIAL REGIME

Nonetheless, several core substantive components of the Supreme Court's opinion in *Freeport-McMoRan* are worthy of substantial critique. Moreover, the subsequent interpretation and application of *Freeport-McMoRan* by the lower federal courts have only amplified the doctrinal distress engendered by that holding.

The contemporary jurisdictional regime of Rule 25 jurisprudence suffers from at least four central flaws. Each consists of multifaceted doctrinal errors, and they all result in a central legal principle that may substantially burden the efficient and equitable resolution of some federal civil actions. These four errors—ironically, yet conveniently—correspond neatly with the four portions of the existing Rule 25 jurisprudence which are principally correct.[159]

---

95 (discussing a related doctrinal deficiency in the Court's articulation of the "indispensable party" limitation in *Freeport-McMoRan* by virtue of the fact that, were a party properly subject to joinder pursuant to Rule 25(c), they would no longer be "indispensable").

153. *See, e.g., Covington Grain*, 638 F.2d at 1364.

154. *See supra* note 151.

155. *See infra* text accompanying note 158 (advancing an analogous interpretation of Rule 25 as applied to both the substitution of indispensable parties and to the overall jurisprudential framework of *Freeport-McMoRan*).

156. *Cf. supra* note 122 (noting and critically examining the contemporary academic consensus that the limitation in *Freeport-McMoRan* is founded on the fact that substitution of an indispensable party will "destroy diversity jurisdiction").

157. *See supra* notes 37-38, 40 (discussing additional, more central, deficiencies).

158. *But see* discussion *infra* Parts III.B-D (addressing several areas in which the Court's analysis was far from insightful).

159. *See* discussion *supra* Part II.

A. The Doctrine: Not Entirely Accurate

First, the jurisdictional doctrine articulated by the Supreme Court in *Freeport-McMoRan,* though properly dispositive of the case before it, was nonetheless both inaccurate and overbroad. The Court concluded that federal diversity jurisdiction persisted, notwithstanding the Rule 25(c) post-filing addition of a nondiverse party, by reference to a central group of cases that, the Court argued, have "consistently held that if jurisdiction exists at the time an action is commenced, such jurisdiction may not be divested by subsequent events."[160] Accordingly, the Court concluded that because the parties originally named in *Freeport-McMoRan* were completely diverse when the action was filed, the subsequent Rule 25(c) substitution did not divest the district court of subject matter jurisdiction.[161] The Court held—in a triumphant conclusion—that "[d]iversity jurisdiction, once established, is not defeated by the addition of a nondiverse party to the action."[162]

The Court, in formulating this assessment, was generally correct as applied to cases arising under Rule 25.[163] Yet, it was entirely wrong as applied to nearly every other form of federal joinder and every other Federal Rule of Civil Procedure.[164] The jurisdictional principle articulated by the Court in *Freeport-McMoRan* may not even properly apply to a large portion of Rule 25 substitutions.[165] In any event, it is clearly inapplicable to the overwhelming majority of cases in which parties are added to a federal civil action, which generally result from the utilization of Rules 15, 19, 20, or 24 rather than Rule 25. The scope and breadth of the Court's central holding is not merely an academic concern, but has also resulted in substantive error by a number of federal courts applying *Freeport-McMoRan.*[166]

1. Cases Originally Filed in Federal Court

Taken at face value, the Court's holding in *Freeport-McMoRan* that "[d]iversity jurisdiction, once established, is not defeated by the addition of a nondiverse party to the action"[167] would substantially alter several preexisting federal jurisdictional principles. For example, federal courts have long held

---

160. Freeport-McMoRan Inc. v. K N Energy, Inc., 498 U.S. 426, 428 (1991) (per curiam); *see also* Wichita R.R. & Light Co. v. Pub. Utils. Comm'n of Kan., 260 U.S. 48, 54 (1922); Clarke v. Mathewson, 37 U.S. 164, 171 (1838); Mollan v. Torrance, 22 U.S. 537 (1824).

161. *See Freeport-McMoRan*, 498 U.S. at 428.

162. *Id.*

163. *But see* discussion *infra* Part II.C (discussing particular Rule 25(c) cases in which neither this statement nor the Court's central holding may properly apply).

164. *See infra* text accompanying note 229.

165. *See* discussion *infra* Part III.C.

166. However, several lower federal courts have been unable to grasp this essential doctrinal failure. *See* cases cited *infra* notes 171, 199.

167. *Freeport-McMoRan*, 498 U.S. at 428.

that the plaintiff's successful post-filing joinder of an additional nondiverse defendant vitiates the original presence of diversity jurisdiction and thereby compels dismissal of the action.[168] This longstanding jurisdictional principle has not only been both codified and extended by Congress in related settings,[169] but also finds additional support in various pre-*Freeport-McMoRan* decisions of the Supreme Court.[170] It is clear—at least prior to the opinion in *Freeport-McMoRan*—that the post-filing addition of a nondiverse party would typically serve to divest the district court of the original federal subject matter jurisdiction that initially existed at the time the action was filed. Thus, the Court's broad dicta and facially unqualified language in *Freeport-McMoRan* may be viewed as raising the question whether the Supreme Court intended, *sub silentio*, to revolutionize the larger jurisdictional structure of federal diversity jurisdiction. Somewhat surprisingly—and clearly erroneously—the answer of several federal district courts has been "yes".

The Supreme Court's clear and unequivocal statement in *Freeport-McMoRan* has not only encouraged litigants to repeatedly advance this holding in an attempt to avoid a jurisdictional dismissal,[171] but has also been relied

---

168. *See, e.g.*, Travelers Indem. Co v. Dingwell, 884 F.2d 629, 636 (1st Cir. 1989) (no diversity jurisdiction over nondiverse parties added pursuant to Rule 19); Lewis v. Lewis, 358 F.2d 495, 502 (9th Cir. 1966) (no diversity jurisdiction over action in which plaintiff amends complaint to join additional nondiverse defendants pursuant to Rule 19); *see also* Acton Co. of Mass. v. Bachman Foods, Inc., 668 F.2d 76, 79-80 (1st Cir. 1982) (extending this principle to the joinder of a nondiverse co-plaintiff as well).

169. *See* 28 U.S.C. § 1367(b) (2000) (no supplemental jurisdiction exists in actions based solely on diversity "over claims by plaintiffs against persons made parties under Rule 14, 19, 20 or 24 . . . or over claims by persons proposed to be joined as plaintiffs under Rule 19 . . . or . . . Rule 24"); *see also infra* text accompanying notes 189-192 (noting implicit jurisdictional assumptions of Rule 19(b)).

170. *See, e.g.*, Owen Equip. & Erection Co. v. Kroger, 437 U.S. 365, 376 (1978) (holding that no diversity jurisdiction exists over a claim by plaintiff against a nondiverse third party defendant added pursuant to Rule 14).

171. *See, e.g.*, Cobb v. Delta Exports, Inc., 186 F.3d 675, 676-77 (5th Cir. 1999) ("The defendants argue that Supreme Court precedent establishes that diversity, for jurisdictional purposes, is established at the time of removal, and the later joinder of dispensable, non-diverse defendants does not destroy federal jurisdiction."); Stubbs v. Kline, No. 97-2133-JWL, 1998 WL 295598, at *1 (D. Kan May 20, 1998) ("Traftec contends that it should be dismissed from the case because diversity is lacking. Plaintiffs disagree because they believe that '[d]iversity jurisdiction, once established, is not defeated by the addition of a non-diverse party to the action.'" (quoting *Freeport-McMoRan*, 498 U.S. at 428)) (footnote omitted) (alteration in original); Williams v. ADT Sec. Servs., Inc., No. Civ.A. 97-3013, 1998 WL 249223, at *1 (E.D. La. May 18, 1998) ("It seems apparent, therefore, that the joinder of American Security—a non-diverse defendant—destroys diversity jurisdiction, and the case must be remanded to state court. Defendants argue, however, that the Supreme Court's decision in *Freeport-McMoRan* . . . militates a contrary ruling.") (citation omitted); Marshall v. CSX Transp. Co., 916 F. Supp. 1150, 1152 (M.D. Ala. 1995) ("The defendants contend that, for diversity purposes, a recent Supreme Court decision bars the court from considering any defendants that are added to a

upon by several federal courts in the same manner. For example, in *Kerr v. Smith Petroleum Co.*,[172] Judge Jones refused to dismiss a diversity action originally filed in federal court after the plaintiffs amended their complaint to add multiple nondiverse defendants.[173] Although both the plaintiffs and the newly-added defendants each argued that the post-filing joinder of these entities destroyed complete diversity and required dismissal, the district court disagreed.[174] The court held that the Supreme Court's express language in *Freeport-McMoRan* was squarely applicable and conclusively established the continuing persistence of federal subject matter jurisdiction over the complaint notwithstanding the contemporary absence of complete diversity amongst the parties.[175]

The Fifth Circuit in *Whalen v. Carter*[176] similarly found *Freeport-McMoRan* both controlling and dispositive of the jurisdictional significance of the post-filing joinder[177] of an additional nondiverse defendant in both Rule 25

---

lawsuit subsequent to the time the lawsuit is filed.") (citing *Freeport-McMoRan*, 498 U.S. at 428).

As these citations reveal, the jurisdictional language in *Freeport-McMoRan* has been employed by litigants of all kinds, and has been utilized by plaintiffs—in an attempt either to authorize the joinder of an additional party defendant or to avoid a jurisdictional dismissal—and by defendants—in an attempt to avoid the dismissal of a favorable judgment or remand to a perceptually inferior forum. *See also* Ingram v. CSX Transp., Inc., 146 F.3d 858, 861 (11th Cir. 1998) (relying on the jurisdictional holding in *Freeport-McMoRan* for the first time on appeal in an attempt to avoid the compelled appellate dismissal for lack of jurisdiction of summary judgment entered below in favor of the appellee).

172. 889 F. Supp. 892 (E.D. La. 1995).

173. *See id.* at 896.

174. *See id.* at 894, 896. It is obvious why the additional defendants desired a dismissal. By contrast, the plaintiffs filed a similar motion—which in a facially counter-intuitive fashion also argued that subject matter jurisdiction did not exist—in order to obviate the district court's previous entry of summary judgment in favor of one of the original defendants. *See id.* at 893-94. Plaintiffs' counsel wisely sought to take advantage of the beneficial tactical consequences of the post-joinder absence of jurisdiction, a fact that would ordinarily suggest representation by skilled and able counsel. This same attorney, however, also appears to have not only "failed to file a timely opposition" to a prior motion for summary judgment, but also sought a "remand" to state court notwithstanding the fact that the action was originally filed in federal court, thereby making potentially available only a dismissal and not a remand. *See id.* at 893-94 & n.2.

175. *See id.* at 896. Judge Jones should not be entirely faulted for reaching such an erroneous conclusion, as he did so in response to facially unambiguous language of the Supreme Court's opinion in *Freeport-McMoRan* and the Fifth Circuit's opinion in Whalen v. Carter, 954 F.2d 1087 (5th Cir. 1992). *See Kerr*, 889 F. Supp. at 896. Accordingly, the decision in *Kerr* cannot be viewed as completely unprincipled given these facially binding although, in truth, entirely distinguishable precedential dictates.

176. 954 F.2d 1087 (5th Cir. 1992).

177. It is somewhat unclear from the Fifth Circuit's opinion precisely how the nondiverse defendant, PHC & Associates, became a party to the action. In *Whalen*, the plaintiffs initially filed three separate lawsuits—which were thereafter consolidated—at least one of which may have initially named PHC & Associates as a defendant. *See id.* at 1089; *see also id.* at 1094

and in non-substitution cases.[178] Accordingly, the court applied *Freeport-McMoRan* to resolve the disputed jurisdictional addition of a nondiverse party who could not otherwise have been permissibly joined.[179] The jurisdictional doctrine articulated in *Freeport-McMoRan*—especially alongside the expansive and facially unlimited language—has thus resulted in the repeated application of *Freeport-McMoRan* to authorize the Rule 20 (and other non-Rule 25) joinder of nondiverse parties in cases originally filed in federal court.[180]

The problem with such holdings is that they are almost certainly wrongly decided. The Supreme Court likely did not intend to entirely abrogate, *ex cathedra*, the traditional jurisprudential assumptions on which the continued assertion of federal diversity jurisdiction had previously been based.[181] Similarly, the Court would be unlikely to do so—even if substantively inclined—through a one-sentence reference in a per curiam opinion that arose in and specifically responded to the peculiar complexities of substitution pursuant to Rule 25(c).[182] Accordingly, even if only as a matter of divining the Court's jurisprudential intent, the language of *Freeport-McMoRan* appears wholly inapplicable to the joinder of additional nondiverse defendants pursuant to provisions other than Rule 25.[183]

The judicial extension of *Freeport-McMoRan* exemplified by the holdings of *Kerr* and *Whalen* appears untenable for two additional extrinsic reasons as well. Such a broad application of *Freeport-McMoRan* would initially appear

---

(noting the initial presence of federal question jurisdiction over these actions prior to the district court's dismissal of the originally articulated claims under RICO). Nonetheless, the Fifth Circuit expressly indicated that "PHC & Associates was added as a defendant after the suit was filed," and accordingly applied the rule articulated in *Freeport-McMoRan*. *Id.* at 1095-96. Hence, it appears that the nondiverse defendant was added pursuant to a post-filing amendment.

178. It is clear, notwithstanding the potential presence of some other ambiguities, that *Whalen* did not involve joinder or substitution pursuant to Rule 25, and the Fifth Circuit's opinion nowhere even mentions this Rule.

179. *See id.* at 1095-96 (quoting exhaustively from and applying *Freeport-McMoRan*, holding in clear reliance thereupon both that diversity jurisdiction is only assessed at the time the action is filed as well as that the addition of a dispensable nondiverse party at no time divests the district court of federal subject matter jurisdiction). Although the Fifth Circuit ultimately concluded that PHC & Associates was in fact an indispensable party and hence could not permissibly be joined in the action, the fact that the Fifth Circuit dispositively applied the "indispensable party" exception of *Freeport-McMoRan* only solidifies the fact that the court viewed the Supreme Court's opinion as applicable to both Rule 25 and non-Rule 25 joinder. *See id.* at 1096. *But see infra* text accompanying notes 184-95 (arguing that *Freeport-McMoRan* does not, in fact, so apply).

180. *See* cases cited *supra* note 171; *see also* discussion *infra* Part III.A.2 (discussing additional cases applying *Freeport-McMoRan* to cases removed to—as opposed to initially filed in—federal court, alongside additional jurisprudential problems arising therefrom).

181. *See supra* text accompanying notes 167-70.

182. *See* discussion *supra* Parts I.B, II.A-B.

183. *See also infra* text accompanying notes 186-95 (discussing additional support for such a conclusion).

inconsistent with and preempted by 28 U.S.C. § 1367(b). Section 1367(b) precludes the assertion of supplemental federal jurisdiction over claims by plaintiffs against persons made parties pursuant to Rules 14, 19, 20, and 24, and also over claims by persons who become plaintiffs under Rules 19 and 24.[184] The application of *Freeport-McMoRan* to the Rule 14, 19, 20, or 24 joinder of a nondiverse party would accordingly appear to authorize what Congress has expressly prohibited; namely, the assertion of ancillary federal jurisdiction over such claims.[185] The Supreme Court cannot typically engage in such a course of conduct,[186] nor should the intent to establish such a holding be presumed lightly.[187] Accordingly, the dictates of 28 U.S.C. § 1367(b) strongly

---

184. *See* 28 U.S.C. § 1367(b) (2000).

185. Section 1367(b) admittedly does not have an expressly preclusive effect, as it only provides particularized exceptions to the supplemental jurisdiction otherwise authorized by § 1367(b). *Id.* It would not be doctrinally unreasonable to assert that the Supreme Court maintains the power to authorize the particularized assertion of common law ancillary jurisdiction by federal courts—at least for diversity actions in which original jurisdiction initially existed at the time of filing pursuant to 28 U.S.C. § 1332—in situations beyond those expressly articulated in 28 U.S.C. § 1367. *Cf.* Hankins v. Finnel, 964 F.2d 853, 860 (8th Cir. 1992) (upholding similar continuing ancillary jurisdiction to conduct postjudgment proceedings, including those to enforce any resulting judgment); Curry v. Del Priore, 941 F.2d 730, 731 (9th Cir. 1991) (upholding the continuing common law inherent power of federal courts to adjudicate collateral matters, e.g., fee disputes, that arise from cases that are originally within the jurisdiction of the federal courts).

Section 1367(b) was nonetheless likely intended to sufficiently cover the matter, and thereby be the—at least *relatively*—exclusive means through which ancillary federal jurisdiction could thereafter be permissibly exercised in those situations to which the statute might facially apply. *Cf.* Shaw v. Meridian Oil, Inc., No. 2:95 CV 1300, 1996 WL 521411, at *1-2 (W.D. La. June 11, 1996) (holding, albeit implicitly, that even in cases in which *Freeport-McMoRan* applies, § 1367 nonetheless provides the exclusive means through which federal jurisdiction over claims not within the original jurisdiction of the federal court may permissibly be maintained). At the very least, when § 1367(b) evidences a deliberate Congressional intent to withhold federal jurisdiction from particular types of claims—claims by plaintiff against parties joined pursuant to Rule 20—it would seem improper for the Court to authorize the exercise of common law ancillary jurisdiction over precisely such contentions.

186. *See* Casas Office Machines, Inc. v. Mita Copystar Am., Inc., 42 F.3d 668, 674 (1st Cir. 1994) (arguing that the "specific legislative directives [of 28 U.S.C. § 1367(b) necessarily] override the general principles" articulated by *Freeport-McMoRan* and relying on this doctrine to limit the application of *Freeport-McMoRan* to the substitution of particular parties for fictitiously named defendants); *see also* Massachusetts v. Andrus, 594 F.2d 872, 887 (1st Cir. 1979) ("Federal courts are courts of limited jurisdiction, and . . . may exercise only the authority granted to them by Congress."); *see generally* Owen Equip. & Erection Co. v. Kroger, 437 U.S. 365, 374 (1978) ("The limits upon federal jurisdiction, whether imposed by the Constitution or by Congress, must be neither disregarded or evaded.").

187. *See* Cobb v. Delta Exports, Inc., 186 F.3d 675, 680 (5th Cir. 1999) ("Courts, of course, may overrule statutes on constitutional grounds, but the limits of diversity jurisdiction are determined purely by statute. Hence, the Court must defer to [statutory dictates], and in reading *Freeport-McMoRan*, we should assume the Court did so.") (citation omitted); *see also*

militate against the validity of the subsequent application of *Freeport-McMoRan* by various federal courts to the joinder of parties added pursuant to provisions other than Rule 25.[188]

The extension of the core *Freeport-McMoRan* holding beyond the particularized context of Rule 25 also conflicts with the substantive dictates of Rule 19. Rule 19 compels the addition of particular ("necessary") parties[189] and establishes a balancing test in the event that one or more of these entities cannot permissibly be joined.[190] If *Freeport-McMoRan* was applicable to joinder pursuant to Rules 14, 19, 20, and 24, very few, if any, parties could not be joined for lack of subject matter jurisdiction—a situation expressly contemplated by Rule 19[191]—since the Court's holding in *Freeport-McMoRan* would allow even entirely nondiverse parties to permissibly be joined.[192] The Court's conclusion that its holding applied only to dispensable entities[193] would mean that indispensable parties could not be joined, and hence would remain subject to the inquiry expressly anticipated by Rule 19. Nevertheless, the application of any such exception would be entirely tautological, as a party would be indispensable only if it could not permissibly be joined, but could permissibly be joined (under *Freeport-McMoRan*) so long as it was not indispensable.[194] As such, Rule 19 appears facially inconsistent with the doctrinal regime provoked by the contemporary application of *Freeport-McMoRan* beyond Rule 25. Particularly when combined with the limitations imposed by both precedent and 28 U.S.C. § 1367, this fact strongly weighs

---

id. at 680-81 (restricting the application of *Freeport-McMoRan* to particular removal cases on the basis of this general jurisprudential principal).

188. *See* cases cited *supra* note 186. *See also infra* text accompanying notes 249-51 (distinguishing the application of Rule 25 as applied to 28 U.S.C. § 1367 on additional bases as well); *infra* text accompanying notes 324-25 (providing additional support for such a limited application of *Freeport-McMoRan*).

189. *See* FED. R. CIV. P. 19(a).

190. *See* FED. R. CIV. P. 19(b).

191. *See* FED. R. CIV. P. 19(a) (articulating procedures that apply in situations in which a person exists "who is subject to service of process and whose joinder will not deprive the court of jurisdiction over the subject matter of the action," thereby recognizing that the joinder of particular parties could in fact divest the court of jurisdiction, yielding a result incompatible, at least in part, with the expansive application of *Freeport-McMoRan*).

192. There may continue to be some exceptions to the general ability of parties to permissibly be joined without divesting the court of subject matter jurisdiction, e.g., the addition of immune or other exceptional parties, but the general superstructure on which Rule 19 is based would remain largely inconsistent with an expansive application of *Freeport-McMoRan* that would permit the vast majority of parties subject to a Rule 19 inquiry to be joined pursuant to the Court's holding.

193. Freeport-McMoRan Inc. v. K N Energy, Inc., 498 U.S. 426, 428 (1991) (per curiam).

194. *See also infra* text accompanying note 296 (advancing an additional related doctrinal critique as a result of the resulting circularity of any expansive application of the *Freeport-McMoRan* rule).

against such a jurisprudential extension in cases initially filed in federal court.[195]

2. Cases Removed to Federal Court

Several courts have not only expansively applied the holding in *Freeport-McMoRan* beyond the context of Rule 25 in original federal actions, but also to actions subsequently removed to a federal forum. Litigants have repeatedly cited the broad jurisdictional language in *Freeport-McMoRan*[196] in an attempt to avoid an otherwise compelled remand to state court[197] by noting that the Court expressly stated that the post-filing joinder of a nondiverse defendant did not divest the district court of federal subject matter jurisdiction.[198] Several federal courts have found such arguments persuasive and, accordingly, allowed the Rule 15, 19, 20, or 24 joinder of a nondiverse defendant while simultaneously refusing to remand the resulting action to state court despite the resulting absence of complete diversity amongst the parties.[199]

As applied to cases subsequently removed to federal court, not only do the same doctrinal concerns discussed in relation to cases originally filed in federal court[200] equally weigh against the permissible application of *Freeport-McMoRan* to joinder conducted pursuant to provisions other than Rule 25, but such holdings are also unsupportable for an additional and substantial reason. As a central part of the special statutory regulations that govern actions removed to federal court,[201] Congress has expressly provided in 28 U.S.C. § 1447(e) that "[i]f after removal the plaintiff seeks to join additional defendants whose joinder would destroy subject matter jurisdiction, the court may deny

---

195. *See supra* notes 186-87.

196. See *Freeport-McMoRan*, 498 U.S. at 428 ("We have consistently held that if jurisdiction exists at the time an action is commenced, such jurisdiction may not be divested by subsequent events. . . . Diversity jurisdiction, once established, is not defeated by the addition of a nondiverse party to the action.").

197. *See* 28 U.S.C. § 1447(c) (2000) ("If at any time before final judgment it appears that the district court lacks subject matter jurisdiction, the case shall be remanded."); *id.* § 1447(e) ("If after removal the plaintiff seeks to join additional defendants whose joinder would destroy subject matter jurisdiction, the court may deny joinder, or permit joinder and remand the action to the State court.").

198. *See, e.g.*, cases cited *supra* note 171.

199. *See* Sharp v. KMart Corp., 991 F. Supp. 519, 523-25 (M.D. La. 1998) (discussing several district court cases that have authorized the non-Rule 25 joinder of nondiverse parties on the basis of *Freeport-McMoRan*); *see also* Bienaime v. Kitzman, No. CIV. A. 00-284, 2000 WL 381932, at *1 n.17 (E.D. La. Apr. 12, 2000) (citing an ostensibly "growing number of" holdings that are contrary to *Freeport-McMoRan* but noting the presence of substantial diametrical authority); Shaw v. Meridian Oil, Inc., No. 2:95 CV 1300, 1996 WL 521411, at *1-2 (W.D. La. June 11, 1996) (applying *Freeport-McMoRan* to adjudicate the validity of plaintiff's Rule 20 joinder of a nondiverse defendant).

200. *See* discussion *supra* Part III.A.1.

201. *See generally* 28 U.S.C. §§ 1445-52 (articulating a plethora of such restrictions).

joinder, or permit joinder and remand the action to the State court."[202] Section 1447(e) facially appears to prohibit a decision—whether pursuant to *Freeport-McMoRan* or any other judicial authority—to join a nondiverse party and to thereafter persist in the federal adjudication of the resulting action.[203] Indeed, it is on the basis of the categorical "plain language" of § 1447(e)[204] that, notwithstanding the contrary authority discussed above, the vast majority of lower federal courts have correctly refused to apply *Freeport-McMoRan* to non-Rule 25 cases that were removed from state court.[205]

Yet the courts that have applied the "plain language" of § 1447(e) have done so on an incomplete and partially erroneous basis. Section 1447(e) expressly addresses the situation in which a plaintiff seeks to join an additional defendant in a removed action. Moreover, the statute clearly indicates the general impermissibility of simultaneous joinder and continuing federal adjudication.[206] These statutory characteristics, however, do not obviate the express prerequisite to the application of § 1447(e) that the proposed joinder of the defendant be one that "would destroy subject matter jurisdiction."[207] As Magistrate Judge Wilson sagely[208] argued in *Shaw v. Meridian Oil, Inc.*:

---

202. *Id.* § 1447(e).

203. *Cf.* Casas Office Machines, Inc. v. Mita Copystar Am., Inc., 42 F.3d 668, 674-75 (1st Cir. 1994) (noting that, pursuant to § 1447(e), a court may alternately (1) deny joinder and thereafter dismiss or retain the action, or (2) permit joinder and remand the case to state court, but "may not, however, do what the court below did here, that is, substitute the nondiverse . . . defendants . . .—thereby defeating federal diversity jurisdiction—and then continue to deal with the merits of the dispute"); *see also infra* text accompanying notes 210-24 (discussing at length the permissible scope of judicial involvement in the application of § 1447 and the generally controlling nature of express statutory commands).

204. *See, e.g.*, Cobb v. Delta Exports, Inc., 186 F.3d 675, 677 (5th Cir. 1999) (arguing that "[t]he plain language of 28 U.S.C. § 1447(e) requires a remand").

205. *See, e.g.*, Bienaime v. Kitzman, No. CIV. A. 00-284, 2000 WL 381932, at *1 n.17 (E.D. La. Apr. 12, 2000) (stating that "[a] growing number of courts have not applied the holding of *Freeport*[*-McMoRan*] when the case involved 28 U.S.C. §1447(e) rather than a Fed.R.Civ.P. [sic] 25(c) transfer of interest"); Stubbs v. Kline, No. 97-2133-JWL, 1998 WL 295598, at *2 (D. Kan. May 20, 1998) (applying 28 U.S.C. § 1447(e) rather than the broad holding of *Freeport-McMoRan*); Williams v. ADT Sec. Servs., Inc., No. CIV.A. 97-3013, 1998 WL 249233, at *2 (E.D. La. May 18, 1998) (limiting *Freeport-McMoRan* to situations "involv[ing] a Rule 25(c) substitution" and remanding the case pursuant to § 1447(e)); Startz v. Tom Martin Constr. Co., No. 93 C 126, 1994 WL 117464, at *2 (N.D. Ill. Mar. 25, 1994) (distinguishing *Freeport-McMoRan*, as the Court failed to "address" § 1447(e)); *see also* Ingram v. CSX Transp., Inc., 146 F.3d 858, 861-62 (11th Cir. 1998) (limiting *Freeport-McMoRan* to substitution under Rule 25(c) and noting the district court's failure to consider the implications of 28 U.S.C. § 1447(e)); Sharp v. KMart Corp., 991 F. Supp. 519, 520, 526-27 (M.D. La. 1998) (applying 28 U.S.C. § 1447(e) rather than the holding in *Freeport-McMoRan*).

206. 28 U.S.C. § 1447(e); *e.g.*, *Cobb*, 186 F.3d at 677; *Bienaime*, 2000 WL 381932, at *1 n.16; *Williams*, 1998 WL 249233, at *1; *Sharp*, 991 F. Supp. at 522.

207. 28 U.S.C. § 1447(e).

208. Magistrate Judge Wilson's analysis in *Shaw* was not entirely doctrinally accurate. *See infra* text accompanying notes 220-24. Nonetheless, his assessment demonstrates a capacity for

> The language of the statute [28 U.S.C. § 1447(e)] limits its application to those situations where joinder would destroy [diversity] jurisdiction. It does not purport to dictate when joinder will destroy jurisdiction. . . . Congress enacted § 1447(e) to expand the options of the district court in those cases where joinder would destroy the court's jurisdiction. Nothing in [the] language or history [of § 1447(e)] suggests that [the statute] was intended to limit the subject matter jurisdiction of the district court.[209]

It is thus insufficient merely to conclude, as existing authority repeatedly states,[210] that the plain language of § 1447(e) manifestly precludes the application of *Freeport-McMoRan* to a case removed from state court. Section 1447(e) appears instead facially to incorporate preexisting substantive jurisdictional authority and merely serves to indicate the procedure to be followed in the event that a requested joinder would divest the court of jurisdiction pursuant to existing law.[211] The proper application of both *Freeport-McMoRan* and § 1447(e) to removed cases is accordingly far from simple. Indeed, were § 1447(e) read to articulate a substantive jurisdictional limitation on joinder and thereby categorically preclude the post-removal addition of nondiverse defendants—as is repeatedly suggested by contemporary post-*Freeport-McMoRan* authority[212]—such an interpretation would be facially applicable to *both* Rule 25 and non-Rule 25 joinder.[213] Such a principle would

---

insightful reasoning, an ability that has been notably absent in the plethora of cases that have arisen on this issue from the great state of Louisiana. *See, e.g., supra* text accompanying notes 172-206 (discussing the cases of *Bienaime*, *Kerr*, *Sharp*, and *Williams*—all from Louisiana—as well as the Fifth Circuit's decisions in *Cobb* and *Whalen*, both of which arose from the decision of a Louisiana federal district court). It is unclear why the federal courts of Louisiana appear so uniquely unwise in their application of Rule 25 or why these courts appear to have articulated a substantial fraction of the relatively scarce published authority on Rule 25 issues.

209. No. 95 CV 1300, 1996 WL 521411, at *2 (W.D. La. June 11, 1996) (citation omitted).

210. *See* cases cited *supra* note 205.

211. *See Shaw*, 1996 WL 521411, at *2; *see also* 28 U.S.C. § 1447(e) (expressly applicable only when joinder "would destroy subject matter jurisdiction" but nowhere attempting to delineate the situations in which such joinder would in fact serve to divest the court of jurisdiction).

212. *See* cases cited *supra* note 205.

213. Joinder of an additional nondiverse defendant pursuant to Rule 25 would, of course, continue to be available notwithstanding the statutory commands of § 1447(e) in situations in which the defendant, rather than the plaintiff, requested the substitution, at least to those courts that found persuasive the D.C. Circuit's analysis in *Burka II*. *See supra* text accompanying notes 85-110 (discussing this holding); *see also supra* Part II.D (critically exploring the doctrinal and equitable validity of this view). The post-removal addition of a nondiverse defendant pursuant to Rule 25 would also not violate the dictates of § 1447(e) were one to find persuasive the D.C. Circuit's untenable distinction between "additions" authorized by Rule 25 as opposed to "joinder" regulated § 1447. *See supra* text accompanying notes 91-103 (criticizing the second holding of *Burka II*). Nevertheless, such purported distinctions would likely fail to preclude the starkly differential and inequitable treatment between cases originally filed in

substantially negate both the breadth and value of the Court's holding in *Freeport-McMoRan*, as application of the *Freeport-McMoRan* rule to removed cases would be precluded by § 1447(e).[214] The resulting view of *Freeport-McMoRan* as applicable only to cases originally filed in federal court, especially when combined with the resulting inequity between removed and non-removed cases, would considerably diminish the effectiveness and legitimacy of the Court's holding.

Fortunately, the doctrinal basis for a view of § 1447(e) as precluding application of *Freeport-McMoRan* to removed actions is insubstantial.[215] Section 1447 generally incorporates, rather than independently establishes, the existing substantive jurisdictional rules. Accordingly, when the Supreme Court held that the substitution of even a nondiverse defendant pursuant to Rule 25(c) does not destroy diversity,[216] this conclusion properly applies to both removed and nonremoved cases, and thereby obviates the application of § 1447(e). Given the Court's holding in *Freeport-McMoRan*, the express prerequisite to application of § 1447(e)—that plaintiff "seek to join additional defendants whose joinder would destroy subject matter jurisdiction"[217]—simply does not apply to substitution pursuant to Rule 25, as joinder in such situations would not in fact result in such jurisdictional consequences. The view that § 1447(e) precludes the Rule 25 substitution or joinder of nondiverse parties in removed cases, notwithstanding *Freeport-McMoRan,* accordingly misinterprets the statute.

The existing removal authorities, however doctrinally misguided, may nonetheless have achieved the correct result, at least as applied to the purported

---

federal court—to which the *Freeport-McMoRan* holding would squarely apply—and to cases removed, in which *Freeport-McMoRan* would ostensibly be "trumped" by § 1447(e) under a *Burka* analysis. *See infra* text accompanying notes 218-24.

214. The Supreme Court's jurisprudential conclusion that Rule 25 substitution does not, as a doctrinal matter, divest the district court of federal subject matter jurisdiction, as well as the policy reasons that support such a result, facially apply equally to both removed and nonremoved cases. There is little reason to believe that the Supreme Court in *Freeport-McMoRan* intended its holding to apply only to a fraction of the federal cases—those initially filed in a federal forum—in which Rule 25 substitution was sought. Indeed, had the Court anticipated that a reversal of the Tenth Circuit would necessitate a substantive and inequitable disparity between removed and nonremoved cases, the Court may perhaps have resolved *Freeport-McMoRan* by reaching the opposite conclusion. Even if the Court had still reversed the Tenth Circuit, the discussion would at least have been more contentious than a slim per curiam opinion suggests.

215. *See Burka II*, 87 F.3d at 482-83 (rejecting, albeit on partially erroneous bases, such a differential application of *Freeport-McMoRan*). *But see* Casas Office Machines, Inc. v. Mita Copystar Am., Inc., 42 F.3d 668, 674-75 (1st Cir. 1994) (holding, in a substantially persuasive opinion, that § 1447(e) nonetheless precludes the application of Rule 25 to the substitution of *fictitiously named* defendants, at least in removed cases).

216. *See* Freeport-McMoRan Inc. v. K N Energy, Inc., 498 U.S. 426, 427-28 (1991) (per curiam).

217. 28 U.S.C. § 1447(e) (2000).

extension of *Freeport-McMoRan* to joinder in removed cases under rules *other* than Rule 25. In such cases, *Freeport-McMoRan* is inapplicable and does not authorize the addition of a nondiverse party. But, this result does not follow from the presently articulated rule that the Supreme Court is categorically precluded from altering the jurisdictional consequences expressly articulated in § 1447(e).[218] Indeed, § 1447(e) actually *incorporates* these extant judicial principles, providing a wholly independent basis for the conclusion that *Freeport-McMoRan* is inapplicable to the joinder of nondiverse defendants obtained through rules other than Rule 25, at least in removal cases.[219] Were the jurisdictional holding in *Freeport-McMoRan* applicable to the entire procedural universe through which the post-filing addition of a nondiverse defendant might be requested—not only to Rule 25, but also to Rules 15, 19, 20, and 24—§ 1447(e) would be effectively meaningless. There would be virtually no situation in which the statutory predicate to this provision—that plaintiff "seek to join additional defendants whose joinder would destroy subject matter jurisdiction"[220]—would apply. Such a holding would conflict not with the "plain language" of § 1447(e),[221] but rather with the central purposes of the statute,[222] the legislative history attendant thereto,[223] as well as core jurisprudential canons of statutory construction.[224]

---

218. *Cf.* Cobb v. Delta Exports, Inc., 186 F.3d 675, 680-81 (5th Cir. 1999) (arguing against the extension of *Freeport-McMoRan* to removal cases on the grounds that the dictates of § 1447(e) overrule any generally contrary jurisdictional doctrine articulated by the Court); *Casas Office Machines*, 42 F.3d at 674 (stating that "[f]ederal courts are courts of limited jurisdiction, and . . . may exercise only the authority granted to them by Congress") (alterations in original).

219. *See supra* text accompanying notes 161-205 (arguing that—wholly apart from § 1447(e)—the holding of *Freeport-McMoRan* is inapplicable beyond the context of Rule 25 for preexisting doctrinal reasons).

220. 28 U.S.C. § 1447(e).

221. *But cf. Cobb*, 186 F.3d at 677 (relying on such purported language to preclude the application of *Freeport-McMoRan*).

222. *See* Sharp v. KMart Corp., 991 F. Supp. 519, 522, 526-27 (M.D. La. 1998) (arguing that such a result would frustrate the clear Congressional purpose in enacting § 1447(e) and the attendant limitation upon permissible post-removal joinder).

223. *See Cobb.*, 186 F.3d at 677 ("The legislative history supports this reading [of *Freeport-McMoRan* as inapplicable to non-Rule 25 removal cases]. In adopting the current version of § 1447(e), Congress rejected a version that would have permitted district courts to join non-diverse defendants and, at their discretion, retain jurisdiction. This history indicates that Congress must not have intended to permit the course of action the district court chose.") (citing David D. Siegel, *Commentary on 1988 Revisions of Section 1447, in* 28 U.S.C.A. § 1447 (West 1994) (citing H.R. REP. NO. 100-889, at 72-73 (1988), *as reprinted in* 1988 U.S.C.C.A.N. 5982, 6032-33)); Williams v. ADT Sec. Servs., Inc., No. Civ.A. 97-3013, 1998 WL 249223, at *2 (E.D. La. May 11, 1998) (finding "[p]articularly persuasive . . . [the] legislative history of § 1447(e)"); *Sharp*, 991 F. Supp. at 526 (arguing that "[t]he legislative history of Section 1447(e) is clear that Congress was careful not to expand the federal court's subject matter jurisdiction in the slightest" and finding significant the fact that "Congress specifically rejected language which would have statutorily allowed a district court to retain jurisdiction even though a non-diverse party was added to the suit after the case had been removed and the parties have established

For these reasons, in addition to those explored above, the principle articulated in *Freeport-McMoRan* and the jurisdictional doctrine relied upon therein[225] are applicable only in the peculiar context of Rule 25. The unqualified language and broad dicta carelessly utilized by the Supreme Court not only "read in a vacuum could vastly change the law in diversity cases where a new party is added who would have otherwise destroyed diversity jurisdiction,"[226] but have also engendered substantial doubt as to the continuing validity of a plethora of pre-*Freeport-McMoRan* jurisdictional precedents.[227] The jurisprudential doctrine articulated by the Supreme Court, while substantially correct as applied to the particular case before it, has unfortunately also resulted in both substantive error by and substantial disputes in a wide variety of post-*Freeport-McMoRan* federal litigation.[228] Notwithstanding the facial breadth of the Court's opinion, this core doctrine is in truth properly applicable only to Rule 25[229] and does not apply, either in removed or nonremoved cases, to substitution or joinder pursuant to provisions other than Rule 25.

---

subject matter jurisdiction pursuant to 28 U.S.C. § 1332"); *cf. Casas Office Machines,* 42 F.3d at 674-75 (relying heavily on analogous legislative history of § 1447(e) to conclude that *Freeport-McMoRan* was inapplicable to the post-filing substitution of particular named for originally filed fictitious defendants in removal cases).

224. *See Sharp.,* 991 F. Supp. at 526 ("If this Court followed *Kerr* and *Shaw*'s interpretation of . . . *Freeport*[*-McMoRan*] in this removal case, then Section 1447(e) would have no meaning."); *see also* Stubbs v. Kline, No. 97-2133-JWL, 1998 WL 295598, at *2 (D. Kan. May 20, 1998) (quoting this finding from *Sharp* with approval).

225. Freeport-McMoRan Inc. v. K N Energy, Inc., 498 U.S. 426, 428 (1991) (per curiam) ("We have consistently held that if jurisdiction exists at the time an action is commenced, such jurisdiction may not be divested by subsequent events. . . . Diversity jurisdiction, once established, is not defeated by the addition of a nondiverse party to the action.").

226. *Sharp,* 991 F. Supp. at 524; *accord Cobb,* 186 F.3d at 680 ("[T]he *Freeport-McMoRan* Court's broad statement that 'diversity jurisdiction, once established, is not defeated by the addition of a nondiverse party to the action,' read in a vacuum, would suggest that the joinder of Waste Management and the city did not destroy diversity. There are good reasons, however, to read this broad statement as *dictum* and to understand *Freeport-McMoRan* as limited to the context of an addition under FED. R. CIV. P. 25." [sic]) (citation omitted).

227. *See Sharp,* 991 F. Supp. at 521 (noting the facially controlling nature of both 28 U.S.C. § 1447(e) and associated precedent but that "there are some cases which suggests [sic] that th[is] . . . analysis is no longer controlling after the United States Supreme Court decision in *Freeport-McMoRan*"); *see also* Bienaime v. Kitzman, No. CIV. A. 00-284, 2000 WL 381932, at *1 n.17 (E.D. La. Apr. 12, 2000) (noting similar jurisprudential disputes engendered by *Freeport-McMoRan*).

228. *See* cases cited *supra* note 205.

229. *See also Cobb,* 186 F.3d at 680 (concluding, albeit on partially erroneous bases, that "[t]here are good reasons . . . to understand *Freeport-McMoRan* as limited to the context of an addition under FED. R. CIV. P. 25" [sic]); Ingram v. CSX Transp., Inc., 146 F.3d 858, 861 (11th Cir. 1998) (concluding, again on somewhat misguided grounds, that "the City's addition was unrelated to [R]ule 25" and that "[a]s a result, we find *Freeport* to be inapplicable").

B. The Reason: Ignoring the Converse

The problem is not merely that the language and central jurisdictional principle in *Freeport-McMoRan* and subsequent interpretive authority are somewhat misguided. The core policy-based justification on which the Supreme Court relied[230] was similarly incomplete. The Court was undoubtedly correct that a jurisdictional rule contrary to the one ultimately articulated in *Freeport-McMoRan* "could well have the effect of deterring normal business transactions during the pendency of what might be lengthy litigation."[231] Litigants, concerned that Rule 25 substitution might destroy diversity, could indeed deter or delay the filing of a motion to substitute, thereby potentially diminishing the equitable and efficient adjudication of the existing action. In addition, such a jurisprudential rule might also produce suboptimal economic arrangements.[232] But, the ability to permissibly substitute even nondiverse parties pursuant to Rule 25(c) could simultaneously be tactically employed to circumvent the complete diversity rule as well as to obtain doctrinally illegitimate procedural benefits. In holding as it ultimately did, the Supreme Court wholly ignored these deleterious practical consequences.Take, for example, a situation in which Mr. Partnoy (a resident of California) and Mrs. Partnoy (a resident of Nevada) are jointly injured in a car accident as a result of the allegedly negligent conduct of Mr. Urquhart, a resident of California.[233] Such a scenario would relatively rarely result in economic inefficiency even if the Supreme Court had concluded that the post-filing substitution of a nondiverse party pursuant to Rule 25 served to divest the district court of diversity jurisdiction. Not only would such a lawsuit by the Partnoys typically be filed in state court anyway (and on behalf of both litigants),[234] but the

---

230. *See supra* text accompanying notes 52-63 (identifying and exploring the fundamental validity of this practical justification).

231. Freeport-McMoRan Inc. v. K N Energy, Inc., 498 U.S. 426, 428-29 (1991) (per curiam).

232. *See supra* text accompanying notes 55-61.

233. This hypothetical assumes (contrary to fact) that Mr. and Mrs. Partnoy are citizens of different states because they reside in a home located exactly on the border between California and Nevada and in which the state line runs between the twin beds on which Mr. and Mrs. Partnoy each spend the vast majority of their respective lives. The hypothetical could easily be altered so that Mr. Partnoy was instead a California corporation, Mrs. Partnoy a Nevada corporate subsidiary, and Mr. Urquhart a California resident responsible for simultaneous economic injury to both Partnoys. This Article employs a less sophisticated (and less common, at least in federal court) example of a simple car crash in the text purely for ease of comprehension.

234. The common law complete diversity rule would typically preclude the initiation of such joint litigation in federal court. *See* 28 U.S.C. § 1332(a); Strawbridge v. Curtiss, 7 U.S. 267 (1806). Whatever tactical advantages might potentially arise from federal adjudication of plaintiff's action would generally not practically justify a decision to litigate in federal court solely on behalf of the diverse Mr. Partnoy. *Cf. supra* text accompanying notes 137-47

relative paucity of situations—apart from the post-filing death or incompetence of one of the parties[235]—in response to which Rule 25 might potentially need to be employed suggest that the policy-based concerns upon which the *Freeport-McMoRan* Court relied are fairly insubstantial as applied to this particular hypothetical.

By contrast, the holding in *Freeport-McMoRan* that Rule 25(c) substitution does not generally divest the district court of original subject matter jurisdiction substantially increases the potential for tactical abuse and attendant inefficiency, inequity, and circumvention of presumptively beneficial doctrinal principles. The Partnoys in such a jurisprudential regime might effectively file a federal action against Mr. Urquhart solely on behalf of the diverse Mrs. Partnoy,[236] have Mrs. Partnoy promptly assign all her rights against Mr. Urquhart to Mr. Partnoy—in a tax-free transaction, no less[237]—thereafter substitute the nondiverse Mr. Partnoy for Mrs. Partnoy pursuant to Rule 25(c), and finally have Mr. Partnoy amend the complaint to assert not only the claims against Mr. Urquhart which were assigned to him, but also his own individual damage claims as well.[238] Such indecorous utilization of *Freeport-McMoRan* is not only far from entirely hypothetical,[239] but may also be expected to increase as the knowledge of both *Freeport-McMoRan* and the attendant possibility of procedural manipulation arising therefrom becomes more diffuse.[240] Accordingly, jurisdictionally permissible Rule 25(c) substitution raises the very real potential for tactical abuse—the jurisprudential flip side of

---

(discussing the potentially abusive means through which the Partnoys might attempt to obtain joint federal adjudication of their claims notwithstanding the complete diversity rule).

235. *See* FED. R. CIV. P. 25(a) (death); 25(b) (incompetence).

236. The "indispensable party" limitation articulated in *Freeport-McMoRan* would not generally preclude such a course of conduct. *See supra* note 139 (noting that joint tortfeasors are not typically deemed to be indispensable parties); *see* cases cited *infra* note 285.

237. *See* 26 I.R.C. § 2523(a) (2000) (permitting spouses to make unlimited tax-free gifts to one another).

238. This is not to say that Mr. Partnoy would automatically be able to successfully employ such a strategy. Indeed there are a variety of procedural means through which the district court *might*—were it to both recognize and choose to employ them—legitimately preclude such course of conduct. Substitution pursuant to Rule 25(c), at least as applied by *Freeport-McMoRan*, merely creates the potential for abusive transactions. Just as a contrary holding in *Freeport-McMoRan* would not have resulted in harmful economic consequences in every case, the actual holding does not necessarily engender actual abuse in every case.

239. *See, e.g.*, GTE Wireless Co. v. Cellexis Int'l. Inc., 341 F.3d 1, 3-4 (1st Cir. 2003) (describing a situation in which the original corporate plaintiff was dissolved and its claims assigned to proposed Rule 25 successor shortly after the action was filed, in situation in which assignee might be entitled to seek damages while original plaintiff had admittedly been able to seek only injunctive relief).

240. This Article may, ironically, perhaps contribute to such detrimental consequences, at least if more readers employ or find persuasive Parts II and III, which identify the problems of the status quo, rather than Part IV, which suggests various solutions to these problems.

the policy-based justifications upon which the *Freeport-McMoRan* Court relied.

The deleterious consequences necessarily created by the Court's holding are not limited to those that arise from tactically abusive Rule 25(c) substitution employed by plaintiffs; rather, they may ensue equally from the pernicious utilization of the *Freeport-McMoRan* rule by defendants. Take, for example, a federal action in which the plaintiff has chosen to litigate solely against Defendant Alpha, having deliberately omitted the nondiverse Defendant Beta from this action principally for strategic reasons: Plaintiff could obtain full relief against Alpha without the disadvantageous substantive,[241] litigation-centered,[242] or jury-related[243] consequences of participation by Beta. Beta in such a scenario would ordinarily be jurisdictionally,[244] and perhaps otherwise,[245] precluded from intervening in the action, and Alpha could not typically obtain a Rule 19(b) dismissal based upon the absence of the dispensable Beta.[246] Absent *Freeport-McMoRan*, plaintiff would ordinarily be able to litigate such a federal action against Alpha alone if she were to decide to do so.[247]

After the Supreme Court's holding, however, Beta may well be able to join successfully as a defendant in the action. Beta could persuade Alpha to transfer

---

241. *Cf.* discussion *infra* Part III.C (discussing situations in which substitution would give rise to a successor defendant that was able to assert additional defenses to the action, possessed a heightened claim for costs or fees, could separately assert individual claims for affirmative relief against the plaintiff, or some combination of these scenarios).

242. Beta may, for example, uniquely possess the financial resources, attitudinal disposition, and ability to tolerate risk required for an aggressive defense of the particular litigation at issue, while Alpha may lack one or all of these attributes and hence be a far more attractive defendant against which the plaintiff would prefer to solely litigate.

243. Alpha may, for example, be a particularly unsympathetic defendant, whereas the participation of Beta may reveal to the jury a perceptually distinct set of facts in which expansive recovery by the plaintiff would be far less likely.

244. *See* 28 U.S.C. 1367(b) (2000) (rejecting supplemental jurisdiction over claims by plaintiff against intervenor defendants added pursuant to Rule 24); Lumber Ins. Co. v. Allen, 892 F. Supp. 31, 33 (D.N.H. 1993) (holding that § 1367(b) precluded a nondiverse entity from intervening as defendant).

245. FED. R. CIV. P. 24(a) (authorizing intervention as a matter of right only when the interests of the proposed intervenor are not "adequately represented by existing parties"). *See, e.g.*, Daggett v. Comm'n on Governmental Ethics & Election Practices, 172 F.3d 104, 111-13 (1st Cir. 1999) (discussing largely proper district court refusal to permit the proposed intervention of defendant—either permissively or as of right—in light of adequacy of representation of existing parties).

246. *See supra* text accompanying notes 130-32 & note 236.

247. Plaintiff could, of course, alternately choose to litigate against both Alpha and Beta in state court rather than solely against Alpha in federal court. The deliberate provision of such an option to the plaintiff arises from the intentional creation of both the complete diversity rule and the substantive dictates of Rule 19.

the relevant underlying interests to Beta,[248] for full value and thereafter obtain inclusion as a substituted or additionally joined party defendant, pursuant to Rule 25. By employing Rule 25 and *Freeport-McMoRan* rather than Rule 19 or 24, Alpha and Beta may successfully obtain the otherwise impermissible joinder of Beta, thereby not only effectively circumventing both the complete diversity rule[249] and the statutory dictates of 28 U.S.C. § 1367(b),[250] but also securing the deliberately unavailable[251] strategic and substantive legal advantages of Beta's participation.

Such a result would not only classically be viewed as injurious and unintended, but would also entail substantial inequity. Defendants would be able to selectively obtain joint federal adjudication of the action notwithstanding the absence of a corresponding ability by the plaintiff to compel such a resolution.[252] The Supreme Court has typically been exceptionally reluctant to construe statutory and Rule-based provisions with jurisdictional import in a manner that could potentially result in such imbalance. For example, in a context roughly analogous to Rule 25, the Supreme Court in *Hoffman v. Blaski*[253] interpreted 28 U.S.C. § 1404(a)[254] to prohibit a defendant from obtaining a transfer of venue to a forum in which the

---

248. Alpha would presumably be more than happy to engage in such a transaction, particularly in light of the pending litigation against it, Beta's willingness to take responsibility for the potential liability, and Alpha's weakness relative to his opponent. Indeed, a transfer of interest from Alpha to Beta in such situations would entail economic synergy; accordingly, were Rule 25 substitution available, Beta might be able to obtain the underlying assets at a discount and to offer Alpha a market premium. All of this, of course, would operate to the detriment of the plaintiff, from whom the economic benefits of the transaction are necessarily obtained. *Cf. infra* Part III.A.1 (discussing related doctrinal issues arising from the potentially collusive creation of federal subject matter jurisdiction, as well as the (generally inadequate, at least as applied to Rule 25) prohibition on such conduct contained in 28 U.S.C. § 1359).

249. *See supra* note 234.

250. *See* 28 U.S.C. § 1367(b) (2000) (divesting statutory supplemental jurisdiction "over claims by plaintiffs against persons made parties under Rule 14, 19, 20, or 24"); *see also supra* note 257; *infra* notes 338-39 (discussing additional substantive components of Rule 25 as applied to § 1367).

251. *See supra* note 247 (noting the conscious jurisprudential and congressional decision to deprive parties like Beta of such benefits).

252. Of course, after *Freeport-McMoRan*, in situations in which the plaintiff has a co-plaintiff and hence herself might be able to creatively employ Rule 25, the rule would be subject to potential categorical abuse by both plaintiffs and defendants alike. *See supra* text accompanying notes 233-38. This generalized "mutuality of inequity" nonetheless does not result in particularized equality, as in any given case, the jurisdictional abuses engendered by Rule 25 typically could be applied only by one side, rather than mutually. *Cf.* Fetterly v. Paskett, 744 F. Supp. 966, 974 (D. Idaho 1990) (mentioning "the old and very well used saying that 'two wrongs don't make a right'").

253. 363 U.S. 335 (1960).

254. See 28 U.S.C. § 1404(a) (2000) ("For the convenience of parties and witnesses, in the interest of justice, a district court may transfer any civil action to any other district or division where it might have been brought.").

action "might have been brought"[255] because the sole basis for the transfer was that the defendant's deliberate post-filing conduct constituted an express or implied waiver of an otherwise existing jurisdictional objection.[256] The Court held that such tactical post-filing activities were insufficient to authorize the requested change in forum.[257] The Supreme Court's equitable analysis in *Hoffman* would perhaps equally apply to the situation presented to the Court thirty years later in *Freeport-McMoRan*.[258] The Court reasoned:

> The thesis urged by petitioners would . . . inject gross discrimination. That thesis, if adopted, would empower a District Court, upon a finding of convenience, to transfer an action to any district desired by the *defendants* and in which they were willing to waive their statutory defenses as to venue and jurisdiction over their persons . . . . Conversely, that thesis would not permit the court, upon motion of the *plaintiffs* and a like showing of convenience, to transfer the action to the same district, without the consent and waiver of venue and personal jurisdiction defenses by the defendants. Nothing in § 1404(a) . . . suggests such a unilateral objective and we should not, under the guise of interpretation, ascribe to Congress any such discriminatory purpose.[259]

By contrast, the Supreme Court in *Freeport-McMoRan*—albeit likely unknowingly[260]—interpreted Rule 25 to produce precisely such potentially inequitable consequences. Accordingly, the central, policy-based foundation of the Court's holding may have achieved the objectives expressly identified, but did so at a substantial—and heretofore unacknowledged—practical and jurisprudential cost.

C. The Extension: Magnifying the Trouble

Subsequent extension of the Court's holding by lower post-*Freeport-McMoRan* federal courts[261] has only amplified the harmful doctrinal and

---

255. *Id.*

256. *See Hoffman*, 363 U.S. at 342-44.

257. The Court's language was potentially applicable far beyond the context of § 1404(a), holding that the jurisdictional status of the action is not properly "made to depend . . . upon the wish of or waiver of the defendant." *Id.* at 343.

258. *Compare id.* at 342 (noting defendant's argument—ultimately rejected by the Court—that "in the interim between the bringing of the action and the filing of a motion to transfer it, the defendants may move their residence to, or, if corporations, may begin the transaction of business in, some other district, and, if such is done . . . empower the District Court to transfer the action") *with* Freeport-McMoRan Inc. v. K N Energy, Inc., 498 U.S. 426, 428 (1991) (per curiam) (supporting its holding by asserting an analogy to the post-filing change in residence of the parties).

259. *Hoffman*, 363 U.S. at 344.

260. *See supra* note 214.

261. *See supra* Part II.C and text accompanying notes 171-205.

pragmatic consequences of this opinion. These various authorities have contributed an abundance of additional substantive errors to contemporary Rule 25 and federal jurisdictional jurisprudence,[262] not the least of which was the episodic extension of *Freeport-McMoRan* far beyond the confines of Rule 25(c).[263]

The most doctrinally and practically significant of these many judicial holdings remains *Burka II*.[264] The opinion of the D.C. Circuit, notwithstanding its many substantive and analytical deficiencies,[265] contains by no means the worst reasoning of the many decisions that have arisen in response to the Supreme Court's ambiguous opinion in *Freeport-McMoRan*.[266] *Burka II* is centrally significant in the subsequent extensions of *Freeport-McMoRan* in light of the D.C. Circuit's conclusion regarding the availability of both supplemental jurisdiction in Rule 25 cases as well as the permissible assertion of additional affirmative claims for relief.

The *Burka II* court was confronted with a previously unaddressed, but fairly common,[267] situation in which a party joined pursuant to Rule 25 asserted additional claims against an existing adverse party.[268] The absence of complete diversity of citizenship between the substituted entity and the proposed adverse party required that jurisdiction over these state law claims be obtained, if at all, pursuant to the supplemental jurisdiction of the district court.[269] The D.C. Circuit, without substantial analysis, concluded that such jurisdiction existed over these additional claims.[270] The application and extension of *Freeport-McMoRan* in *Burka II* not only heightened the potential for circumvention of both the complete diversity rule and the intended functions of 28 U.S.C. § 1367, but also exponentially multiplied the possibility of tactical abuse of Rule 25. The Supreme Court's more limited holding in *Freeport-McMoRan* undoubtedly gave rise to particular doctrinal and pragmatic difficulties.[271] Still, the D.C. Circuit's opinion substantially upped the ante. After *Freeport-McMoRan*, defendants who successfully joined an action pursuant to Rule 25

---

262. *See supra* text accompanying notes 171-205.

263. *See supra* text accompanying notes 180-82.

264. *Burka II*, 87 F.3d 478 (D.C. Cir. 1996); *see also* discussion *supra* Part II.C (analyzing this opinion at some length).

265. *See, e.g., supra* text accompanying notes 75-109 (critically exploring, *inter alia*, the articulation in *Burka II* of a purported distinction between "addition" versus "joinder" as applied to the interaction between Rule 25(c) and 28 U.S.C. § 1447, as well as various purported substantive characteristics of the Supreme Court's holding in *Freeport-McMoRan*).

266. *Cf. supra* text accompanying notes 171-205 (critiquing some terrible district court opinions, particularly those of district courts located in Louisiana).

267. *See id.*

268. *See Burka II*, 87 F.3d at 479-80 (noting American University's assertion upon being added as a defendant of various counterclaims against the plaintiffs based upon conduct allegedly performed during the pendency of the action).

269. *See* 28 U.S.C. § 1367 (2000); *Burka II*, 87 F.3d at 480.

270. *See Burka II*, 87 F.3d at 480 n.4.

271. *See* discussion *supra* Parts II.A, III.A; *supra* text accompanying notes 120-42.

could at least facially persist in the continuing defense of the same claim that was already tasked for federal adjudication,[272] potentially resulting in appreciable tactical and substantive legal benefits.[273] Nonetheless, the incentive to utilize Rule 25(c) in a manipulative fashion would be, at least as compared to the situation in *Burka II*, relatively low.

By contrast, the authorized articulation of additional affirmative claims for relief provides a substantial impetus to engage in precisely such abusive conduct. According to *Burka II*, defendants can permissibly procure the federal adjudication of a claim against a nondiverse plaintiff through the relatively simple process of obtaining a transfer of the underlying assets in a bona fide assignment from the existing defendant.[274] Given the potentially substantial benefits of such an outcome,[275] the D.C. Circuit's opinion generates the possibility that Rule 25 could practically function not merely as intended—a procedural device designed to authorize the continued and efficient resolution of the *existing* dispute[276]—but also to allow the addition of multiple state law contentions over which there is jurisprudential distaste for federal adjudication, and thereby operate in a far more practically and strategically dispositive manner than planned.

Nevertheless, the very real potential for Machiavellian utilization and doctrinal manipulation of Rule 25, *Burka II*, and *Freeport-McMoRan* by defendants pales in comparison to the availability of such abusive tactics available to plaintiffs. Given the contemporary economic emphasis on operative synergies,[277] the scenario described in the Alpha and Beta

---

272. *Freeport-McMoRan* neither squarely foreclosed nor manifestly authorized the articulation of additional supplemental claims by the substituted party. Yet, the mere fact that the *Freeport-McMoRan* Court was not presented with a situation in which the substituted party desired to do anything other than persist in the articulation of the existing claim would presumably have given subsequent lower courts some pause before interpreting this decision to allow the post-substitution articulation of additional state law claims between a flatly nondiverse plaintiff and defendant. *But see Burka II*, 87 F.3d at 480 n.4 (authorizing such claims, consistent with *Freeport-McMoRan*, in a single, two-sentence footnote).

273. *See, e.g.*, discussion *supra* Parts II.B, III.B (describing enhanced recovery of costs, damages, and various additional extralegal strategic benefits).

274. Of course, this is not to suggest that such a transfer will in fact be available in each and every situation in which one is required. Moreover, in order to authorize the permissible exercise of supplemental jurisdiction, the claim to be asserted by the defendant must arise from the same Article III case or controversy as the plaintiff's preexisting contentions. *See* 28 U.S.C. § 1367(a). Particularly in light of the fairly expansive jurisprudential interpretation of the latter requirement, there may be a wide variety of situations in which a transfer of interest pursuant to Rule 25(c) would be both practically available as well as strategically beneficial. *See* United Mine Workers v. Gibbs, 383 U.S. 715, 725 (1966); *see also, e.g., Burka II*, 87 F.3d at 480 (involving precisely such a scenario).

275. *See* discussion *supra* Parts II.B, III.B.

276. *See supra* text accompanying notes 43-44.

277. *See* Total Containment, Inc. v. Intelpro Corp., Nos. 99-1059, 99-1060, 1999 WL 717946, at **3 (Fed. Cir. Sept. 15, 1999); *see also supra* text accompanying notes 241-52

hypothetical[278] may in fact often arise, particularly in intellectual property cases.[279] Even in more routine litigation, there exists a wide number of situations in which the plaintiff has a substantial incentive to circumvent the complete diversity rule and obtain joint federal resolution of the claims against each of the allegedly responsible defendants.[280] The D.C. Circuit's opinion in *Burka II* not only expands *Freeport-McMoRan* to potentially authorize precisely such federal adjudication, but similarly allows the strategic Rule 25 joinder of multiple plaintiffs[281] notwithstanding the absence of complete diversity amongst the adverse parties. It is one thing for the Supreme Court in *Freeport-McMoRan* to hold that hypothetical Mr. Partnoy could permissibly substitute for his deceased, or incompetent, wife pursuant to Rule 25 without compelling a dismissal of the pending action for lack of federal jurisdiction. But, it is doctrinally and pragmatically quite another for the D.C. Circuit in *Burka II* to hold that Mr. Partnoy could not only continue to litigate his wife's claims on her own behalf, but also assert his own separate claims for relief. The permissible expansion of the existing action—especially to include additional contentions that could not otherwise be resolved in the federal judiciary—both substantially and qualitatively augments the already appreciable implications of the Supreme Court's holding in *Freeport-McMoRan* and the resulting jurisprudential regime.

D. The Caveat: Simultaneously Overbroad and Underinclusive

One final practical and doctrinal problem worthy of mention stems not from the subsequent extension of *Freeport-McMoRan* by the lower federal courts, but instead arises from the Supreme Court's opinion. The Supreme Court was fundamentally correct in the articulation of a Rule 25 exception for "indispensable parties"[282] Still, that central qualification was simultaneously overbroad, underinclusive, and substantively inaccurate.

The underinclusive nature of the *Freeport-McMoRan* Court's indispensable party exception is readily apparent. Each of the previous hypotheticals[283] involved a situation in which the nondiverse party to be added pursuant to Rule

---

(discussing such a contemporary business focus and the resulting rise of merger, acquisition, and divestment activities).

278. *See supra* text accompanying notes 241-52.

279. *See, e.g.*, GTE Wireless Co. v. Cellexis Int'l, Inc., 341 F.3d 1, 3 (1st Cir. 2003); *Total Containment*, 1999 WL 717946, at **1-3.

280. *See* discussion *supra* Part III.B.

281. *See supra* text accompanying notes 233-40.

282. *See supra* text accompanying notes 115-21 (supporting the propriety of such a core limitation, at least as applied to the substitution of particular nondiverse parties).

283. *See supra* text accompanying notes 138-47, 233-40, 241-52 (advancing the Monster and Tiny, Partnoy and Urquhart, Alpha and Beta, and various other distinct scenarios under which Rule 25 substitution might tactically be abusively employed).

25 was entirely dispensable,[284] and accordingly the Supreme Court's doctrinal limitation is insufficient to constrain the potential abuse necessarily engendered by the central holding. The relatively narrow set of circumstances in which an absent party will, in fact, be indispensable[285] reflects the manifest inability of the sole limitation in *Freeport-McMoRan* to properly restrict the jurisdictional reach of that holding to situations in which a substitution pursuant to Rule 25 should properly be allowed.

The indispensable party exception is also somewhat substantively misguided. Litigants undoubtedly should not generally be permitted to vitiate a dismissal otherwise compelled by Rule 19(b)—thereby altering a preexisting, dispositive defense—simply through the expedient of substitution pursuant to Rule 25.[286] Such a result, however sufficient, is neither necessary nor accurately reflective of the factors that should control whether or not a particular Rule 25(c) substitution should be permissible. The core focus in any Rule 25 inquiry should not concern whether a party was initially legally indispensable, but rather should focus on the resulting scope and jurisdictional effect of such substitution on the existing action.[287]

The propriety of Rule 25 substitution pursuant to *Freeport-McMoRan* rests conclusively on the fine balance between the legitimate need for such joinder and the level of abuse potentially associated therewith.[288] Accordingly, it is on this basis—not the tangentially relevant Rule 19 status of the proposed successor—that a Rule 25 substitution should largely stand or fall.[289] The *Freeport-McMoRan* Court's focus on Rule 19 in articulating the sole jurisdictional limitation accordingly appears, if nothing else, largely to miss the point.

---

284. *See* cases cited *supra* notes 171-205 and accompanying text.

285. *See, e.g.*, Temple v. Synthes Corp., 498 U.S. 5, 8 (1990) (closely related joint tortfeasors held not indispensable to action involving only one such defendant); *see also* Stabilisierungsfonds Fur Wein v. Kaiser Stuhl Wire Distribs. Party Ltd., 647 F.2d 200, 207 (D.C. Cir. 1981).

286. *See* discussion *supra* Part II.B (defending the validity of such a result on both Rule 25 and extrinsic doctrinal grounds); *but see* discussion *supra* Part II.D (asserting limitations on such a rule); *infra* Part IV.B (arguing that substitution even of an initially indispensable party should occasionally be permitted notwithstanding the Court's holding in *Freeport-McMoRan*).

287. *See* discussion *infra* Part IV.B (advancing such an alternative jurisprudential regime). Again, this is not to say that an assessment as to the dispensable or indispensable status of the party to be joined pursuant to Rule 25(c) should not be employed, as such an inquiry may be a dispositively sufficient basis upon which to deny such substitution even when a doctrinally unnecessary precondition.

288. *See* discussion *supra* Part II.B, D.

289. *See id.*; *see also* discussion *infra* Part IV.B (noting this central component of *Freeport-McMoRan* and accordingly exploring the doctrinal and pragmatic reliance upon the exercise of particularized discretion as a means of resolving the disputed scope and application of Rule 25).

The Court's indispensable party limitation is also doctrinally obtuse in its origins and substantive content. This obscure[290] caveat—likely derived from the roughly analogous context of common law Rule 24 intervention[291]—facially has little applicability to either Rule 25 or the jurisdictional application to 28 U.S.C. § 1367. Indeed, the facile importation of doctrinal concerns relating to whether or not a particular party was "dispensable" when the action was initially filed might appear largely irrelevant to the permissible *jurisdictional* status of such a party upon substitution. After all, the absence of a party "to be joined" pursuant to Rule 19 does not categorically deprive the district court of jurisdiction. Furthermore, the procedural deficiency of a claim for failure to satisfy the requirements of Rule 19, unlike the absence of subject matter jurisdiction,[292] may not only occasionally be waived,[293] but also does not necessarily compel the dismissal of an action filed without the initial joinder of an indispensable party.[294] Accordingly, there is not an automatic correlation between the jurisdictional condition of a party added pursuant to Rule 25 and the distinct status of a party under Rule 19.

The migration in *Freeport-McMoRan* of a doctrine applicable to Rule 19 to the differential context of Rule 25 may also be internally inconsistent. A party is "indispensable" under Rule 19 only if it cannot permissibly be joined.[295] However, *Freeport-McMoRan* facially authorizes the joinder of entities even over which the court would ordinarily lack jurisdiction. The definition of a Rule 25 indispensable party is accordingly somewhat circular, as a particular party would qualify as indispensable only if he could not be joined, and yet could permissibly be joined under *Freeport-McMoRan* so long as he was not

290. *See supra* text accompanying notes 115-21 (describing the backhanded genesis and articulation of this doctrine).

291. *See* discussion *supra* Part II.D (speculating as to and critiquing the validity of such origins).

292. *See* Sosna v. Iowa, 419 U.S. 393, 398 (1975); Franzel v. Kerr Mfg. Co., 959 F.2d 628, 629-30 (6th Cir. 1992).

293. *See, e.g.*, Citibank, N.A. v. Oxford Props. & Fin. Ltd., 688 F.2d 1259, 1262 n.4 (9th Cir. 1992) (holding that the failure to join a party as required by Rule 19(a) may be waived); B.F. Goodrich Co. v. Goodyear Tire & Rubber Co., No. 86 C 3145, 1987 WL 6295, at *1 n.1 (N.D. Ill. Feb. 5, 1987) (relying upon its conclusion to permit defendants to "waive[] any claim that BFG failed to join an indispensable party"); *see also* Judwin Props., Inc. v. U.S. Fire Ins. Co, 973 F.2d 432, 434-35 (5th Cir. 1992) (holding that a plaintiff generally cannot appeal an adverse ruling for his own failure to join an indispensable party pursuant to Rule 19(b) notwithstanding the ability to appeal similarly adverse judgments for lack of subject matter jurisdiction).

294. *See* Provident Tradesmens Bank & Trust Co. v. Patterson, 390 U.S. 102, 114 (1968); *see also* FED. R. CIV. P. 19(a) (permitting continuation of a federal action where the post-filing joinder of the indispensable party was accomplished).

295. FED. R. CIV. P. 19(b).

indispensable.[296] Accordingly, the jurisprudential confluence of jurisdiction under Rule 25 and indispensable parties under Rule 19 is far from clear.

Of course, there are several similarities between the failure to join a Rule 19 indispensable party and a dismissal for lack of subject matter jurisdiction,[297] and both courts and commentators have often imparted virtual jurisdictional significance to the initial failure to join an indispensable party.[298] Presumptively it is for these reasons that the Supreme Court in *Freeport-McMoRan*, subsequent federal authorities, and academic commentators alike have made the inferential leap from the initial absence of a particular indispensable party and the jurisdictional inability to subsequently join that entity pursuant to Rule 25.[299]

It may perhaps be jurisprudentially and academically sufficient merely to demonstrate the doctrinal inadequacy of such a conclusion.[300] Regardless, apart from any purely intellectual interest in the matter, the questionable transmutation of the "indispensable party" principles to the jurisdictional significance of Rule 25 substitutions has at least one critical substantive implication. The relevant period during which the successor party may not permissibly be indispensable pursuant to the *Freeport-McMoRan* limitation is expressly the time at which the action is filed.[301] By contrast, the classic Rule

---

296. *See supra* text accompanying note 194 (discussing this potential deficiency in more substantial detail).

297. *See, e.g.*, FED. R. CIV. P. 12(h)(2) (permitting such an objection to be raised at any time); CP Nat'l Corp. v. Bonneville Power Admin., 928 F.2d 905, 911-12 (9th Cir. 1991) (allowing the nonjoinder of an indispensable party to be raised both for the first time on appeal and by the Court of Appeals *sua sponte*).

298. *See, e.g.*, Schuckman v. Rubenstein, 164 F.2d 952, 957 (6th Cir. 1947) (The "failure of the district court to acquire jurisdiction over indispensable parties to an action deprives the court of jurisdiction to proceed in the matter and render a judgment."); Note, *Indispensable Parties in the Federal Courts*, 65 HARV. L. REV. 1050, 1050 (1952) ("The absence of an 'indispensible' party is fatal to the maintenance of a suit."); Note, *Indispensable Parties Under the Federal Rules of Civil Procedure*, 56 YALE L.J. 1088, 1089 (1947) ("Interested persons are called 'indispensible' and must be joined if the judgment will affect their rights or liabilities even if the consequence is dismissal of the case.") (citation omitted).

299. *See, e.g.*, Cobb v. Delta Exports, Inc., 186 F.3d 675, 679 (5th Cir. 1999) (interpreting *Freeport-McMoRan* to hold that "if the party that was added [pursuant to Rule 25] had been indispensable when suit was filed, the addition of the non-diverse party would have defeated diversity jurisdiction"); Sharp v. KMart Corp., 991 F. Supp. 519, 523 (M.D. La. 1998) ("The [*Freeport-McMoran*] Court found that if the party who has been added was indispensable at the time the plaintiff filed its complaint, the addition of a non-diverse party would defeat diversity jurisdiction."); *cf.* 7 CHARLES ALAN WRIGHT, ARTHUR R. MILLER & MARY KAY KANE, FEDERAL PRACTICE AND PROCEDURE § 1610 (3d ed. 2001) (arguing, albeit in a different context, that ancillary jurisdiction presupposes the existence of an action that a federal court can adjudicate and should not be used to breathe life into a lawsuit that is not otherwise viable).

300. *See supra* text accompanying notes 290-94 (distinguishing the concepts of Rule 25 substitution and the Rule 19 status of parties).

301. *See* Freeport-McMoRan, Inc v. K N Energy, Inc., 498 U.S. 426, 428 (1991) (per curiam) (noting that the indispensable party inquiry relates to whether the Rule 25 successor

19 inquiry typically focuses not on whether the party was indispensable at the time of filing, but rather whether the party is *currently* indispensable.[302]

As a result, the substantive content of the *Freeport-McMoRan* limitation is simultaneously over- and underinclusive. Clearly, there are situations in which a party is initially indispensable and yet, at some point during the pendency of the action, no longer maintains that status.[303] A Rule 19 dismissal would be entirely unavailable in such settings.[304] Similarly, regardless of whether or not that party was originally indispensable, no categorical doctrinal reason explains why such a party could not presently be properly substituted pursuant to Rule 25. Situations may also exist in which a party was not initially indispensable and yet thereafter became so. Permissible Rule 25 substitution in such cases would substantially alter the otherwise dispositive existing defenses and hence would appear improper,[305] but a court's express focus on the status of the party at the time the action was filed[306] would nonetheless preclude an application of the *Freeport-McMoRan* limitation to such cases. Accordingly, the caveat in *Freeport-McMoRan* may seem properly focused on the initial propriety of the litigation, but in fact articulates a doctrine that is imprecise and fails to capture the essential bases on which a Rule 25 substitution should properly be denied.

The indispensable party exception of *Freeport-McMoRan*, however, should not necessarily be applied in the same temporal fashion as the inquiry required by Rule 19.[307] To the degree that a Rule 19 focus on the potentially dispensable status of a party should be incorporated in the context of Rule 25 substitution,[308] the appropriate time at which the condition of the absent party should be assessed is immediately prior to the transfer of interest, or other

---

possessed such a status "at the time the complaint was filed"); *see also supra* text accompanying notes 171-205 (citing numerous post-*Freeport-McMoRan* federal courts identically so holding).

302. *See, e.g.*, Provident Tradesmens Bank & Trust Co. v. Patterson, 390 U.S. 102, 114 (1968); *see also* FED. R. CIV. P. 19 (assessing the Rule 19 status of an absent entity according to the particular facts that exist at the time the motion to dismiss is filed).

303. For example, this situation may arise because a party has changed his residence and hence may properly be joined without destroying diversity, has now conducted sufficient post-filing activity in the forum state, or has engaged in an effective waiver, so to authorize the exercise of personal jurisdiction, or simply no longer has the type of "interest" or other factors relevant to the litigation required to qualify as an indispensable party pursuant to Rule 19.

304. *See* discussion *supra* Part II.D.

305. *See id.* (noting the doctrinal impermissibility of such a result).

306. *See id.*

307. For example, it would be entirely improper to assess the status of the party proposed to be joined pursuant to Rule 25 based upon the circumstances that existed at the time the relevant motion was filed, which is precisely how Rule 19 is applied, as the underlying transfer of interest that gave rise to the motion might—at that point—result in the proposed successor being indispensable, and yet there is no reason why such parties should not permissibly be added to the action. *See* discussion *supra* Part II.D.

308. *Compare* discussion *supra* Part II.D (arguing for such an application in particular, limited circumstances) *with* discussion *supra* Part III.A (rejecting the generalized application of such a principle).

event, that gave rise to the request for Rule 25 substitution. Such an inquiry, rather than the one articulated in *Freeport-McMoRan*, would properly assess the propriety of the proposed substitution and, at least in part, the risk that Rule 25 could be employed to avoid an otherwise compelled Rule 19 dismissal.

Accordingly, the Supreme Court's opinion in *Freeport-McMoRan* was simultaneously centrally well-founded and yet critically misguided on a variety of different issues. Rule 25 can and should be employed in particular cases notwithstanding the resulting lack of complete diversity between the parties. Nonetheless, the existing jurisprudential regime that has arisen from the Court's opinion in *Freeport-McMoRan* and subsequent expansion by the lower federal courts is far from perfect and creates a broad assortment of substantial doctrinal and pragmatic disorders.

IV. IMPROVING THE EXISTING RULE 25 JURISDICTIONAL REGIME

Given the plethora of difficulties with the contemporary regime of Rule 25 jurisprudence, the question naturally arises: "What is to be done?"[309]

Simply identifying the respective problem is sufficient, at least as applied to several matters, thereby informing both litigants and, ultimately, the responsible judicial decisionmakers of the substantive errors of the various holdings. The variety of mistakes in the application of 28 U.S.C. § 1447(e) to both Rule 25 and non-Rule 25 removal cases,[310] the D.C. Circuit's untenable substantive distinction between the "addition" as opposed to the "joinder" of a party,[311] the categorical inapplicability of *Freeport-McMoRan* to joinder conducted pursuant to provisions other than Rule 25—and the inconsistency between the contrary lower court holdings and common law precedent, 28 U.S.C. §§ 1367, 1447(e), and Rule 19[312]—and the temporal and doctrinal inaccuracy of the *Freeport-McMoRan* "indispensable party" exception[313] are each self-contained errors that may be resolved by mere repudiation of the underlying doctrine.[314] Such a response would dramatically improve the existing jurisprudential regime.

Nonetheless, there remain a variety of more intractable problems with contemporary Rule 25 jurisprudence, which include, but are by no means

---

309. *Cf.* VLADMIR I. LENIN, *What Is to Be Done?*, *in* THE LENIN ANTHOLOGY 13, 50 (Robert C. Tucker ed. 1975) (asking "what must be done to bring political knowledge to the workers").

310. *See supra* text accompanying notes 201-205 (discussing these issues).

311. *See supra* text accompanying notes 264-76 (critiquing *Burka II*).

312. *See supra* text accompanying notes 180-95 (critically examining this authority).

313. *See supra* text accompanying notes 301-08 (exploring the deficiencies of the contemporary limitation and advancing an alternative doctrinal regime).

314. Because few of these doctrines, notwithstanding their pervasive nature, arise from directly controlling authority, such as a holding of the Supreme Court or—apart from *Burka II*—a squarely applicable holding from a Court of Appeals, federal courts are relatively free to "do the right thing" in these areas of Rule 25 and related jurisprudence.

limited to, the central problem of distinguishing between the permissible and impermissible substitution of nondiverse parties.[315] In fact, a variety of different potential solutions may exist. Some of these solutions may successfully resolve only particular problems. Moreover, all solutions may be expected to succeed entirely—if such a result is even possible—only as an integrated whole. Particular cases may also demand the application of particular doctrines, and it may be valuable to separately identify the different possible responses to the problems established by the existing Rule 25 jurisprudence.

A. Possible Categorical Solutions

1. Application of the Statutory Rule Against Collusive Jurisdiction

Initially, the strategic assignment of claims pursuant to Rule 25(c) in an effort to obtain tactical and jurisdictional advantages obviously presents a situation where the statutory rule against the creation of collusive jurisdiction might be properly applied.[316] No federal court has attempted to apply 28 U.S.C. § 1359 to preclude the addition of a party pursuant to Rule 25, but the facial breadth and central purpose of this statutory provision seems entirely consistent with an effort to ensure that Rule 25 is not employed in order to circumvent the complete diversity rule and obtain the benefits of an otherwise unavailable federal forum.

Section 1359, while potentially applicable to particular types of Rule 25(c) substitution, nonetheless cannot realistically be expected to substantially resolve the problems resulting from the jurisdictional holding in *Freeport-McMoRan*. First, 28 U.S.C. § 1359, which expressly applies only to efforts to "*invoke* the jurisdiction of the court,"[317] may be entirely inapplicable to post-filing activities in situations to which Rule 25(c) might apply. After all, federal jurisdiction has already been properly *invoked* by the time a disputed motion to substitute under Rule 25 (for "post-filing" events) is made, and § 1359 may, for this reason alone, be inapposite as applied to efforts merely to extend—rather than *invoke*—an existing federal adjudication. Particularly because post-transfer motions to substitute are generally made in order to permissibly *continue* an existing federal action, § 1359 seems to be a somewhat ill-suited means to address the problems created not by the collusive creation of initial federal jurisdiction, but that arise instead merely from the inclusion of additional parties therein.

---

315. *See supra* text accompanying notes 183-95 (discussing this core component of *Freeport-McMoRan* and associated jurisprudence).

316. *See* 28 U.S.C. § 1359 (2000) ("A district court shall not have jurisdiction of a civil action in which any party, by assignment or otherwise, has been improperly or collusively made or joined to invoke the jurisdiction of such court.").

317. *Id.* (emphasis added).

Second, even if § 1359 was potentially applicable to substitution pursuant to Rule 25, as a practical matter, it is unlikely that the problems associated with the holding in *Freeport-McMoRan* would be substantially resolved through reliance on this statute. Although § 1359 may be dispositive of entirely collusive assignments lacking substantive reason—that have no real purpose whatsoever other than to invoke federal diversity jurisdiction[318]—such a doctrine may have little practical application to Rule 25(c), as the overwhelming majority of Rule 25 transfers accomplish at least some legitimate function.[319] Moreover, not only do the underlying economics of abusive Rule 25(c) transactions typically allow the undertaking of a bona fide transaction,[320] thus obviating the application of § 1359,[321] but the subjective nature of the underlying inquiry[322] would often permit litigants to effectively manufacture facially persuasive rationalizations for the transfer and avoid dismissal.

Section 1359 finally appears substantively misguided—at least as applied to Rule 25—for two additional reasons, each of which would compel non-exclusive reliance thereupon. First, § 1359 is entirely too amorphous and subject to variable interpretation and application to provide a complete solution to the various problems engendered by *Freeport-McMoRan*. The multiple factors and heavily fact-dependent nature of the inquiry compelled by § 1359[323]—especially when combined with the substantive malleability of the statute[324] and the appreciable transaction costs necessarily associated with any such adjudication—not only weighs heavily against dependence upon § 1359, but also indicate the value of any alternative bright-line approach to the underlying jurisdictional problem.

In any event, the potential application of 28 U.S.C. § 1359 largely misses the point. The problem with *Freeport-McMoRan* is not merely that parties may respond with a successful effort to collusively enlarge the scope of federal adjudication, but also that—even in situations in which the underlying transfer of interest is entirely legitimate—the action should *still* not expand to include the resolution of additional state law claims possessed by the nondiverse successor against the existing adverse parties. Collusive or not, the exercise of

---

318. *See* Kramer v. Caribbean Mills, Inc. 394 U.S. 823, 828 (1969).

319. *See, e.g.*, GTE Wireless v. Cellexis Int'l, Inc., 341 F.3d at 1, 3 (1st Cir. 2003) (corporate dissolution); *Burka II*, 87 F.3d 479, 480 (D.C. Cir. 1996) (purchase of underlying property by proposed successor); *see supra* notes 19-20 (citing numerous additional Rule 25(c) cases involving an assignment, merger, or economic divestment).

320. *See supra* text accompanying note 274.

321. *See* Westinghouse Credit Corp. v. Shelton, 645 F.2d 869, 871 (10th Cir. 1981) (holding that the assignment "was neither 'improperly' nor 'collusively' made to permit diversity litigation in federal court"); *cf. Kramer*, 394 U.S. at 828 (applying 28 U.S.C. § 1359 to preclude jurisdiction over a claim assigned for one dollar in an obviously collusive fashion).

322. *See* Dweck v. Japan CBM Corp., 877 F.2d 790, 793 (9th Cir. 1989).

323. *See* Yokeno v. Mafnas, 973 F.2d 803, 809-10 (9th Cir. 1992); *Dweck*, 877 F.2d at 792-93.

324. *See supra* text accompanying notes 318-21.

ancillary jurisdiction over such claims is neither advantageous nor desired by the overarching jurisprudential order.[325] Section 1359 provides no solution to this central problem, regardless of whether, in particularly egregious cases, it might perhaps be effectively employed to preclude a given Rule 25 substitution.

2. Strictly Applying the "Post-Filing" Transfer Requirement

Strict application of the requirement in Rule 25(c) that the underlying transfer of interest that gave rise to the substitution transpire only after the litigation has commenced is yet another potential solution that is facially applicable to the problems developed by the existing regiment of Rule 25 authorities. Such an approach may effectively preclude the abusive application of *Freeport-McMoRan* by particular litigants. Nonetheless, like section 1359, this doctrine is also far from a panacea for the problems associated with the joinder of nondiverse parties pursuant to Rule 25.

Rule 25(c) has no applicability to transfers of interest that have occurred prior to the filing of an action, but rather applies only to transfers of interest that take place during the litigation.[326] Particular Rule 25 substitutions may perhaps violate this basic principle, at least if the underlying doctrine is strictly construed in a deliberate effort to avoid potential jurisdictional abuses. Take, for example, a situation in which both Original (from Delaware) and Successor (from Hawaii) have jointly signed a contract with Performer (from Hawaii), that Performer has breached. Original subsequently files a federal action against Performer, seeking specific performance. Thereafter, Successor acquires the assets of and dissolves Original. This example contains a "transfer of interest" —the underlying merger and transfer of assets—and Rule 25(c) may allow the post-acquisition substitution of Successor for Original.

Nonetheless Performer might attempt—perhaps persuasively—to argue that there is no *relevant* transfer of interest in this scenario, and that Rule 25(c) is inapplicable, given that Successor, as a joint signator, was already entirely interested in both the underlying contract and in Original's effort to obtain specific performance even before the transfer.[327] Some tangential judicial

325. *See supra* text accompanying notes 241-52 (discussing the potential benefits of categorical abrogation of supplemental jurisdiction over such claims).

326. *See* JAMES WM. MOORE, *supra* note 121 § 25.31[3] (3d ed. 2006); *see also* Travelers Indem. Co. v. Dingwell, 884 F.2d 629, 633-34 (1st Cir. 1989) (noting similar limitations on Rule 25(c) motions). Conversely, if a transfer occurs before the commencement of a suit, Rule 17 applies.

327. *Cf.* DeVilliers v. Atlas Corp., 360 F.2d 292, 294 (10th Cir. 1966) (holding that successor company was properly substituted as plaintiff for predecessor when the successor, prior to the transfer of interest, had no connection with the existing action); Freeport-McMoRan Inc. v. K N Energy, Inc., 498 U.S. 426, 428 (1991) (per curiam) (authorizing substitution of nondiverse party but expressly noting that this entity "had no interest whatsoever in the outcome of the litigation until sometime after suit was commenced").

support may exist for such an argument;[328] but, in situations in which both Original and Successor would seek identical relief, there may be no substantial reason to particularly care whether or not substitution is allowed.[329] Assume, however, that Original sought only specific performance but that Successor, upon substitution, intended to assert an alternative claim for damages that could have been—but was not in fact—alleged by Original. The outcome of the Rule 25(c) motion in such a setting might be exceptionally important, and a strict interpretation of the "post-filing transfer of interest" requirement may serve to preclude substitution in such a situation, on the grounds that Successor had no relevant post-filing transfer of interest on which to properly move for substitution.

There is little existing precedent that would squarely support such a holding,[330] but any such rule might also result in the preclusion of substitution in situations in which it would be entirely beneficial; e.g., when both Original and Successor seek the identical relief and state law precludes the continued post-acquisition assertion of a claim by the dissolved Original, but not the persistent Successor.[331] Ironically, the more potentially abusive the Rule 25(c) substitution becomes, e.g., were Successor to have different, independent claims against Performer that it would promptly articulate upon effective substitution, as in the Partnoy and Urquhart example,[332] the less likely that application of the "post-filing transfer of interest" requirement would in fact successfully preclude the substitution. Starkly different claims for relief would facially indicate that the proposed successor did in fact receive something *different* from the originally named party as a result of the transfer and qualify

---

328. *See, e.g., Travelers Indemnity*, 884 F.2d at 634 (agreeing with the district court that joinder was properly denied in part because the proposed plaintiff "had a direct and substantial interest in the declaratory judgment action at the moment the complaint was filed"); *but see Burka II*, 87 F.3d at 480-81 (allowing substitution pursuant to Rule 25(c) notwithstanding the fact that American University, the proposed additional party, had already signed a purchase contract—and was manifestly interested in the outcome of the existing dispute—at the time the action was filed).

329. There might actually be a substantial amount at stake, especially if the post-filing dissolution of Original precluded the continuation of the action on behalf of Original as a matter of state law. *See supra* note 149 (noting that Rule 25(c) is entirely procedural in nature and that the substantive ability of a dissolved corporate entity to continue to litigate a federal action is resolved according to state law); *see also supra* note 150 (noting that the laws of most—but by no means all—states permit most dissolved entities to continue to litigate an action even post-dissolution).

330. *See supra* note 3.

331. The fact that Successor might also be held to be an initially indispensable party in such a scenario, and substitution held to be impermissible in any event pursuant to the Court's opinion in *Freeport-McMoRan*—even if Successor had been relieved post-filing from such a status, e.g., had mutually abrogated the contract with Performer—provides yet another example of the overinclusive nature of the *Freeport-McMoRan* limitation. *See* Delta Fin. Corp. v. Paul D. Comanduras & Assocs., 973 F.2d 301, 303-04 (4th Cir. 1992).

332. *See supra* text accompanying notes 233-40.

as a party subject to legitimate substitution pursuant to Rule 25(c). The strict doctrinal application of the "post-filing transfer of interest" requirement may thus serve to limit the otherwise expansive reach of the existing Rule 25 jurisprudence, but it is hardly the type of effective solution to the problems resulting from *Freeport-McMoRan* which the contemporary doctrine so desperately needs.

B. Discretionary Application of Rule 25

Although the categorical approaches discussed above do not appear to entirely, or even substantially, resolve the many problems associated with the existing Rule 25 regime, there is an easily applied doctrine that, properly employed, could in fact strongly mitigate these pragmatic difficulties. Rule 25 imparts upon federal district courts virtually unlimited discretion to either grant or refuse a particular substitution or joinder requested.[333] District court judges may permissibly decline to authorize substitution in any situation in which such an event would be tactically abusive, circumvent the complete diversity rule, conflict with 28 U.S.C. § 1367, or otherwise bring about deleterious social or litigation-related consequences. Such a purposeful application and interpretation of the substantive dictates of Rule 25 would be entirely consistent with both the text and purpose of Rule 82, which expressly provides that "[t]hese rules shall not be construed to extend or limit the jurisdiction of the United States district court or the venue of actions therein."[334] Federal courts—if sufficiently interested and informed[335]—could accordingly accomplish the precise objectives advanced at length through the deliberate and targeted exercise of their acceptable discretion.

The problem, of course, with such an individualized approach is that it necessarily relies upon the differential knowledge, skill, interest, and attention of district court judges. Wholesale reliance upon the wisdom of an imperfect, substantially overworked federal judiciary is simply far from the bright-line rule that is increasingly the ideal of contemporary adjudication. The individualized exercise of judicial discretion would almost assuredly leave an appreciable number of cases in which the provisions of *Freeport-McMoRan* and Rule 25(c) would in fact continue to be abused, and would also result in substantive inequity, as different federal judges reached different conclusions in relevantly identical cases. Finally, reliance upon discretion would result in particular situations in which district court judges would erroneously deny substitution, notwithstanding the damaging consequences, resulting in precisely the type of situation that the Supreme Court in *Freeport-McMoRan* expressly intended to avoid.

---

333. *See supra* text accompanying notes 171-205 (citing numerous authorities so holding).

334. FED. R. CIV. P. 82.

335. This Article is obviously designed, in substantial part, to contribute to the fulfillment of such a necessary predicate.

The substantial discretion imparted by Rule 25, however, should not be ignored or overlooked. The circumspect application of both *Freeport-McMoRan* and Rule 25 is undeniably an important component of any effort to constrain the abusive utilization of contemporary Rule 25 jurisprudence. Moreover, there are particular bright-line rules that district courts should perhaps—at least presumptively—employ as an integral part of their considered exercise of discretion. The destructive consequences of *Freeport-McMoRan* and *Burka II,* for example, would be substantially (though not entirely) mitigated if district courts were to generally condition the approval of a proposed Rule 25 substitution on an agreement by the successor to articulate only the claims to which the original party to the action was initially entitled, and therefore essentially agree not to invoke the supplemental jurisdiction of the Court[336]—notwithstanding their otherwise extant authority to do so pursuant to *Burka II*—in an attempt to expand the action beyond its original confines.[337] The informed and intelligent exercise of district court discretion to deny Rule 25 substitution in particular abusive cases—especially if combined with the general imposition of conditions in any litigation in which substitution is eventually allowed—would at least provide some demonstrable improvement in the existing jurisprudential regime.

C. Regulating the Exercise of Supplemental Jurisdiction

Courts and Congress might participate in the alleviation of the adverse consequences caused by *Freeport-McMoRan* and the D.C. Circuit's subsequent expansion of this opinion in *Burka II* in one additional manner as well; namely, through the deliberately restrictive application of 28 U.S.C. § 1367. Section 1367(b), which does not mention Rule 25, does not restrict the exercise of supplemental jurisdiction over claims by or against parties added pursuant to Rule 25(c),[338] but affirmatively grants broad authority to hear precisely such ancillary allegations.[339] Parties may exponentially expand the scope of the existing action on the basis of § 1367 and circumvent a variety of existing jurisprudential limitations through the expedient of Rule 25 joinder.[340]

---

336. *Cf. infra* text accompanying notes 338-42 (proposing analogous categorical and individualized regimes that might be applied pursuant to 28 U.S.C. § 1367).

337. *Cf. Burka II*, 87 F.3d at 480 (providing a concrete example of the failure of existing district court discretion in light of the district and appellate court approval of the successor party's articulation of additional supplemental state law counterclaims possessed exclusively by the nondiverse successor).

338. *See* 28 U.S.C. § 1367(b) (2000) (divesting district courts of supplemental jurisdiction over claims by or against parties joined pursuant to a wide variety of different Federal Rules of Civil Procedure, but not those added pursuant to the provisions of Rule 25).

339. *See* 28 U.S.C. § 1367(a) (granting supplemental jurisdiction in both diversity and nondiversity cases, so long as the claims arise out of the same Article III case or controversy as the originally filed action).

340. *See supra* text accompanying notes 267-81.

Such an application of Rule 25 raises two potential responses. First, Congress could—and likely should—amend § 1367(b) to expressly include parties added pursuant to Rule 25[341] as amongst those governed by the generalized exception to the authority granted to § 1367(a). Congress would need to carefully ensure that the abrogation of supplemental jurisdiction did not negate the Court's fundamentally beneficial holding in *Freeport-McMoRan*, and also to limit only the exercise of *additional* adverse claims by the nondiverse party, as opposed to the continuation of the contentions originally at issue. But, careful drafting should prove entirely possible for the skilled draftsperson.[342]

Similarly, in the event that § 1367 is replaced—as may eventually transpire—with an alternative approach that returns to the common law distinction between ancillary and pendant jurisdiction, those charged with the articulation of the substantive principles should not follow their predecessor's example and entirely ignore Rule 25. Instead, they should expressly address the variety of jurisdictional complexities established by *Freeport-McMoRan* and the subsequent jurisprudential regime.

Second, district court judges can and should employ their discretion pursuant to the provisions of Rule 25, but district courts should generally decline to exercise supplemental jurisdiction over additionally expansive claims alleged by a successor party. Such a procedure is independently authorized by 28 U.S.C. § 1367(c).[343] Undoubtedly situations occur in which the supplemental state law claims alleged by the nondiverse party added pursuant to Rule 25 either raise a novel or complex issue of state law or substantially predominate over the claims over which the court has original jurisdiction, and hence properly authorize the court to decline to exercise supplemental jurisdiction pursuant to §§ 1367(c)(1) and 1367(c)(2). More categorically, the damaging doctrinal and pragmatic consequences often associated with the exercise of supplemental jurisdiction over parties added pursuant to Rule 25(c)[344] may present precisely the type of "exceptional" and "compelling reasons," especially when combined with the intended function of Rule 82, that would allow a similar response pursuant to 28 U.S.C. § 1367(c)(4).

CONCLUSION

Rule 25 and its precedent currently authorize the broad articulation of supplemental state law claims between nondiverse parties in federal court. Regardless of the particular means through which this is accomplished, the

341. Congress might simultaneously elect to similarly include parties joined as plaintiffs pursuant to Rule 20, *see supra* text accompanying notes 180-82.

342. *See supra* text accompanying note 325 (articulating such a proposed approach).

343. 28 U.S.C. § 1367(c) (expressly granting district courts discretion to decline to exercise supplemental jurisdiction in certain situations).

344. *See supra* text accompanying notes 330-32 (describing such consequences at some length).

mitigation and elimination of such contentions should be a high priority for the next evolution of Rule 25 jurisprudence.

The contemporary scope and breadth of Rule 25—and particularly the expansive exercise of supplemental jurisdiction—was almost assuredly never intended. The incremental, superficial, and largely uninformed development of Rule 25 jurisprudence has engendered a situation whose deleterious consequences are far greater than their individual components. It cries out for intelligent revision.

THE ROBERTS COURT: YEAR 1

LORI A. RINGHAND*

INTRODUCTION

The 2005 Term of the U.S. Supreme Court is of extraordinary interest to Court observers. For the first time in eleven years, the Court's Term commenced without Chief Justice Rehnquist and ended without Justice O'Connor. The departure of these two Justices marked the end of one of the longest natural court periods (a period when there are no personnel changes on the court) in history. The departure of Justice O'Connor, long the ideological center of the Court in many areas, also marked a potential shift in the Court's jurisprudence.

This paper examines the Court's seminal 2005 Term "by the numbers": by empirically analyzing what the Court actually did, rather than substantively evaluating what it should have or could have done. In doing so, I consider the number of decisions that the Court issued, the issues that those decisions involved, the ideological direction of the decisions, and the vote margins by which the Court issued those decisions. I also examine the voting records of the new Justices, Chief Justice Roberts and Justice Alito, and the ideological direction of their votes, the issue areas in which they were most active, the number of concurring and dissenting opinions they each wrote, and the Justices with whom they most frequently aligned themselves.

This analysis yields several surprises. First, despite the addition of two new and presumably conservative Justices, the ideological direction of the Court's decisions did not change dramatically. Second, the Court was, at least for this one Term, more unified than it had been in the recent past, issuing a higher percentage of unanimous decisions and a lower percentage of five to four decisions than it did in 2004. Finally, while Chief Justice Roberts and Justice Alito both, as generally expected, aligned themselves most frequently with the Court's conservative Justices, a careful examination of their voting records shows that these Justices actually aligned themselves more closely with each other than with the Court's established conservatives. Chief Justice Roberts also showed a surprisingly high rate of agreement with some of the Court's more liberal Justices, although not in the Term's most controversial decisions.

---

* Associate Professor of Law, University of Kentucky College of Law; J.D. University of Wisconsin Law School; B.C.L. Oxford University. Thanks are owed Professor Neil Siegal for his thoughtful comments on an early draft of this paper, and to Jonathan Milby and Nathan Goodrich for their extraordinary and prompt research assistance.

EXPLANATION OF THE DATA

This paper relies on the datasets and coding methodology developed by Harold Spaeth for use in the U.S. Supreme Court Databases ("Spaeth Databases").[1] The Spaeth Databases include information about all U.S. Supreme Court decisions issued from 1953 to 2004.[2] When making comparisons between the 2005 Term and earlier Terms, data about the earlier Terms was culled from either the Spaeth Original Supreme Court Database or the Spaeth Justice-Centered Database, as updated and amended for my earlier work.[3] I coded the Court's 2005 Term myself, following Spaeth's coding methodology.[4] My analysis throughout this paper includes only those cases in which the Court heard oral argument and issued a formal, full opinion. This includes per curium decisions and plurality decisions but does not include memorandum opinions and decrees.[5]

---

1. Harold J. Spaeth, U.S. Supreme Court Databases, http://www.as.uky.edu/polisci/ulmerproject/sctdata.htm (last visited Sept. 18, 2006) [hereinafter Spaeth Databases]. The National Science Foundation supported the creation of the Spaeth Databases.

2. *Id.*

3. The Spaeth Justice-Centered dataset was revised and updated by me for use in a forthcoming article. *See*Spaeth Databases, *supra* note 1. I updated and revised the Spaeth Original Supreme Court Database for use in another article. *See* Lori A. Ringhand, *The Rehnquist Court: A 'By the Numbers" Retrospective*, U. PA. J. CONST. L. (forthcoming Spring 2007) [hereinafter *The Rehnquist Court*]. For the revised Justice-Centered Dataset see Lori A. Ringhand, Judicial Activism Dataset, http://www.uky.edu/Law/faculty/ringhand.html. For the revised Supreme Court Dataset, see Lori A. Ringhand, Rehnquist Retrospective Dataset, http://www.uky.edu/Law/ faculty/ringhand.html.

4. For the dataset I created to analyze the 2005 Term, see Lori A. Ringhand, Roberts: Year 1 Databasess, http://www.uky.edu/Law/faculty/ringhand.html.

5. I also made certain substantive changes to the Spaeth Databases. I did this when my examination of the underlying cases indicated that a coding error had occurred or that the coding protocols used by Spaeth were plainly substantively inappropriate for this project. The only systemic substantive change made to the Spaeth Databases involved a coding choice made by Spaeth that resulted in some cases arising under the Eleventh Amendment or the Fourteenth Amendment being coded as raising questions of state-level judicial review, even though the Court in these cases actually considered the constitutionality of a federal statute. I changed this coding to reflect that the legally relevant decision in these cases involved the constitutionality of the federal, not the state, statute. A full explanation of the issue area groupings can be found in the Original U.S. Supreme Court Judicial Database Codebook. HAROLD J. SPAETH, THE ORIGINAL UNITED STATES SUPREME COURT JUDICIAL DATABASE 42-52 (2006), http://www.as.uky.edu/ polisci/ulmerproject/allcourt_codebook.pdf [hereinafter Spaeth Codebook]. For a full list of the coding changes made to the publicly available Spaeth Databases, see Lori A. Ringhand, Changes to Spaeth Datasets: Explanatory Document, http://www.uky.edu/Law/faculty/ringhand.html.

I. THE 2005 TERM: DECISIONS OF THE COURT

A. The cases: issue areas, ideological direction, and majority opinion writers

In the 2005 Term, the Roberts Court heard oral argument and issued full, formal opinions in sixty-nine cases.[6] The issues raised in these cases were quite similar to those addressed by the Court in its 2004 Term. As shown below, the 2005 Roberts Court and the 2004 Rehnquist Court both issued most of their decisions in two issue areas, criminal procedure and economic activity. Each court also issued a significant percentage of its decisions in the civil rights and judicial power areas:[7]

Table 1
2004 Term & 2005 Term—Cases by Issue Area

	2004 Term	2005 Term
Criminal Procedure	27.0 %	29.0 %
Economic Activity	18.9	23.2
Civil Rights	16.2	13.0
Judicial Power	14.9	13.0
First Amendment	8.1	8.7
Due Process	8.1	2.9
Federalism	4.1	4.3
Miscellaneous	2.7	2.9
Privacy	0.0	1.4
Attorneys	0.0	1.4

6. For a list of these cases, with citations, vote margins, issue areas, ideological direction, and each Justice's vote, see Appendix A. I did not code for all of Spaeth's variables, opting instead to code only for the variables used herein.

7. For a full explanation of the issue area groupings, see the Spaeth Codebook, *supra* note 5, at 45-56. Each case is coded as involving only one issue. Following Spaeth, I have coded issue areas from a public policy, rather than a formally legal, standpoint. *See id.* The "criminal procedure," "civil rights," "First Amendment," and "federal taxation" issue areas are self–explanatory. "Due process" cases include only non-criminal cases raising due process issues. "Economic activity" cases include cases raising commercial and business issues, including tort actions and non-civil rights based employer/employee disputes. "Privacy" cases include abortion and sexual orientation cases; I also included the 2005 "right to die" case in this category. "Attorneys" cases include only those cases addressing the unique role of attorneys within the legal system, including fee cases. "Judicial power" cases include questions of federal court jurisdiction or authority; "federalism" cases include cases arising under Section Five of the Fourteenth Amendment, Eleventh Amendment cases, and Commerce Clause cases.

The ideological direction of the decisions issued in the two Terms did vary, however. Spaeth codes decisions and opinions as ideologically "conservative" or "liberal."[8] The ideological coding, while complex, generally follows expected, current political preferences: decisions in favor of an individual asserting a constitutional right are coded as liberal, excepting "reverse race discrimination" and campaign finance cases in which a decision in favor of the constitutional claimant is coded as conservative.[9] Decisions in favor of the exercise of federal government or judicial power are coded as liberal, as are decisions in favor of unions, injured plaintiffs, and debtors.[10] Decisions favoring states' rights, compelled arbitration, and property owners (in Takings Clause cases) are coded as conservative.[11]

Of the sixty-nine decisions issued by the Roberts Court in the 2005 Term, 56.5% (thirty-nine cases) generated a conservative outcome, and 39% (twenty-seven cases) generated a liberal outcome. Four percent (three cases) were ideologically uncodable.[12] This record is notably different than that of the Court's 2004 Term. In that year, the Court heard oral argument and issued formal opinions in seventy-four cases, of which only 46% (thirty-four cases) generated conservative outcomes, almost 53% (thirty-nine cases) generated liberal outcomes, and one was uncodable.

The Roberts Court thus appears to have generated notably more conservative opinions than did its predecessor. This is, however, somewhat misleading. While the Court in its 2005 Term did issue a higher percentage of conservative opinions than it did in its 2004 Term, it actually was the 2004 Term that was the ideological anomaly. Since the Rehnquist Court attained its final composition in 1994, when Justice Breyer replaced Justice Blackmun, the Court has issued more conservative than liberal opinions in each Term *except* the 2000 and 2004 Terms. The Court's 2004 Term was, in fact, the Court's most liberal Term since 1994:

---

8. For a full explanation of Spaeth's Coding protocols, which I follow here, see Spaeth Codebook, *supra* note 5, at 52-60.

9. *See id.* at 52-53.

10. *See id.* at 54.

11. *See id.* at 53-54.

12. Cases that do not fit any of Spaeth's ideological codes are considered uncodable, as are cases with such fractured opinions that no single ideological direction can fairly be assigned to the opinion. *See id.* at 52. Three cases were deemed ideologically uncodable here. *See* League of United Latin Am. Citizens v. Perry, 126 S. Ct. 2594, 2626 (2006) (a deeply fractured opinion involving a Texas redistricting effort that rejected a "political gerrymandering" claim while upholding a Voting Rights Act claim that one of the newly created districts depleted the voting rights of Hispanic citizens); eBay, Inc. v. MercExchange, L.L.C. 126 S. Ct. 1837, 1838 (2006) (a unanimous decision involving a business practices patent); Wachovia Bank v. Schmidt, 126 S. Ct. 941, 945 (2006) (an eight to zero decision involving the location of a bank for purposes of diversity jurisdiction).

Table 2 Ideological Outcome of Decisions—1994-2005		
Term	**Conservative**	**Liberal**
1994	54 %	46 %
1995	51	49
1996	59	41
1997	59	41
1998	59	41
1999	51	49
2000	49	51
2001	59	41
2002	55	45
2003	52	48
2004	46	53
2005	56.5	39

As shown above, between the 1994 and 2004 Terms, 53% of the decisions issued by the Supreme Court have been ideologically conservative. The Roberts Court's 2005 Term record of 56.5% conservative opinions is thus only slightly above the Rehnquist Court's performance in recent years.[13]

Given that a majority of the decisions issued in the 2005 Term were ideologically conservative, it is perhaps not surprising that Justices Scalia and Thomas and Chief Justice Roberts (along with Justice Kennedy) authored most of the Term's majority opinions. The number of majority opinions written by each Justice, organized by the issue areas that the opinions involved is as follows:

13. The decisions of the modern Supreme Court have been fairly evenly divided ideologically: since 1953 (when the Spaeth Databases began), 48% of the Supreme Court's formally issued opinions, excluding decrees and memorandum opinions, have been ideologically conservative, and 52% have been ideologically liberal. *See* Spaeth Databases, *supra* note 1. Note, however, that Spaeth's ideological coding is purely relative, meaning that an outcome is "liberal" or "conservative" only relative to the other possible outcome in the case. *See* Spaeth Codebook, *supra* note 5, at 52-60. In other words, these terms are not keyed to any substantive political position or ideology and are, thus, subject to movement based on the types of cases heard by the Court and the positions taken by the parties in those cases. *See id.*

Table 3
Number of Majority Opinions Written by Each Justice—By Issue Area

	Criminal Procedure	Civil Rights	First Amendment	Due Process	Privacy
Per Curiam	1	0	0	0	0
Stevens	1	1	0	0	0
O'Conner	0	1	0	0	0
Scalia	5	1	0	0	0
Kennedy	3	1	1	0	1
Souter	1	1	1	1	0
Thomas	3	1	0	0	0
Ginsburg	0	1	0	0	0
Breyer	2	1	2	0	0
Roberts	2	0	2	1	0
Alito	2	1	0	0	0
Total	20	9	6	2	1

Table 3 (Continued)
Number of Majority Opinions Written by Each Justice—By Issue Area

	Attorneys	Economic Activity	Judicial Power	Federalism	Misc.	Total
Per Curiam	0	0	0	0	0	1
Stevens	0	4	0	1	0	7
O'Conner	0	1	0	0	1	3
Scalia	0	2	0	1	0	9
Kennedy	0	2	0	0	0	8
Souter	0	1	2	0	0	7
Thomas	0	2	1	1	0	8
Ginsburg	0	1	4	0	1	7
Breyer	0	2	0	0	0	7
Roberts	1	1	1	0	0	8
Alito	0	0	1	0	0	4
Total	1	16	9	3	2	69

Also not surprising, given his new position as the presumptive ideological center of the Roberts Court, is the fact that the opinions authored by Justice Kennedy generated almost an identical number of liberal and conservative outcomes:[14]

14. For the ideologically conservative opinions authored by Justice Kennedy, see *Anza v. Ideal Steel Supply Corp.*, 126 S. Ct. 1991, 1998 (2006) (denying a RICO claim for lack of proximate cause); *Garcetti v. Ceballos*, 126 S. Ct. 1951, 1962 (2006) (holding that the First Amendment does not protect statements made by government employees in the course of their employment); *Rice v. Collins*, 126 S. Ct. 969, 973 (2006) (upholding as race neutral a prosecutor's preemptory strike of an African-American juror). For the ideologically liberal opinions he authored, see *Hill v. McDonough*, 126 S. Ct. 2096, 2099 (2006) (allowing a challenge to the method of execution used in death penalty cases); *House v. Bell*, 126 S. Ct. 2064, 2068 (2006) (allowing a convicted criminal to request a new hearing in light of new evidence); *Dolan v. U.S. Postal Serv.*, 126 S. Ct. 1252, 1255 (2006) (holding that the Federal Torts Claims Act did not bar a tort claim against the U.S. Postal Service); *Gonzales v. Oregon*, 126 S. Ct. 904, 925 (2006) (holding that the federal Attorney General did not have statutory authority to block assisted suicides). The uncoded opinion was the Texas redistricting case, *League of United Latin American Citizens v. Perry*, 126 S. Ct. 2594 (2006).

Table 4 Ideological Direction of Decision by Majority Opinion Writer				
Majority Opinion Writer	**Conservative Opinions**	**Liberal Opinions**	**Uncoded Opinions**	**Total**
Per Curiam	0	1	0	1
Stevens	1	6	0	7
Scalia	7	2	0	9
Kennedy	**3**	**4**	**1**	**8**
Souter	5	2	0	7
Thomas	6	1	1	8
Ginsburg	2	4	1	7
Breyer	4	3	0	7
Roberts	6	2	0	8
Alito	2	2	0	4
O'Connor	3	0	0	3
Total	39	27	3	69

B. Judicial unity: vote margins and separate opinions

At his confirmation hearings, Chief Justice Roberts frequently expressed a desire to bring greater unity to the Supreme Court.[15] Consistent with this goal, the Supreme Court during his first term was slightly more unified than it was during Chief Justice Rehnquist's last term. The Roberts Court in the 2005 Term issued a lower percentage of five to four decisions and a higher percentage of unanimous decisions, than did the Rehnquist Court in the 2004 Term:

15. *Confirmation Hearing on the Nomination of John G. Roberts, Jr. to be the Chief Justice of the United States: Hearing Before the S. Comm. on the Judiciary*, 109th Cong. (2005), *available at* 2005 WL 22037049 (F.D.C.H.).

Table 5 Vote Margins in 2004 and 2005 Terms		
Vote Margin	2004	2005
5-2	0%	1.4%
5-3	5.4	4.3
5-4	21.6	14.5
6-2	2.7	2.9
6-3	12.2	11.6
7-1	0.0	2.9
7-2	14.6	8.7
8-0	6.8	17.4
8-1	6.8	2.9
9-0	29.7	33.3

Although the political dynamics on the Supreme Court certainly have shifted since Chief Justice Rehnquist first assumed leadership of the Court in 1986, Chief Justice Roberts has achieved notably greater unity in his inaugural term than Chief Justice Rehnquist did in his:[16]

16. There is, however, at least one interesting similarity between the 1986 and the 2005 Terms. In each of those Terms, seven of the nine Justices sitting were appointed by Republican presidents. In 1986, Chief Justice Rehnquist (appointed by Nixon and elevated by Reagan), Justices Scalia (Reagan), Blackmun (Nixon), Brennan (Eisenhower), O'Connor (Reagan), Powell (Nixon), and Stevens (Ford) were appointed by Republican presidents, while only Justices White (Kennedy) and Marshall (Johnson) were appointed by Democrats. *See* U.S. Supreme Court webpage, Members of the Supreme Court of the United States, http://www.supremecourtus.gov/ about/members.pdf (last visited Sept. 19, 2006). In 2005, Chief Justice Roberts (G.W. Bush), Justices Stevens (Ford), Scalia, Kennedy (Reagan), Souter (G.H.W. Bush), Thomas (G.H.W. Bush), and Alito (G.W. Bush) were appointed by Republican presidents, while only Justices Ginsburg (Clinton) and Breyer (Clinton) were appointed by Democrats. *See id.*

Table 6 Vote Margins in 1986 and 2005 Terms		
Vote Margin	**1986**	**2005**
5-2	0.7 %	1.4%
5-3	1.4	4.3
5-4	28.4	14.5
6-2	2.8	2.9
6-3	22.1	11.6
7-0	1.4	0
7-1	0	2.9
7-2	11	8.7
8-0	2.8	17.4
8-1	4.8	2.9
9-0	24.8	33.3

As shown above, the 1986 Court issued approximately 55% of its opinions with vote margins of six to three or closer, while only approximately 35% of the 2005 opinions were.

Furthermore, it appears that Chief Justice Roberts, as promised in his confirmation hearings, has actively promoted this unity by voting with the majority more often than any other Justice. As shown below, in the 2005 Term Chief Justice Roberts voted with the majority in fifty-eight of the sixty-seven cases in which he took part. Justice Scalia, in contrast, wrote or joined the most concurring opinions, while Justices Stevens and Breyer were the Term's most frequent dissenters. Of the Justices that served the full Term, Justice Stevens joined the fewest majority opinions:[17]

17. These categories are mutually exclusive, meaning that a Justice joining or writing a concurring opinion is not also counted as voting with the majority in that case.

Table 7 Vote Direction by Justice—2005 Term[18]			
	Wrote or joined Dissent	Wrote or joined Majority	Wrote or joined Concurrence
Stevens	17	43	8
Scalia	7	50	11
Kennedy	8	53	7
Souter	12	52	4
Thomas	13	50	4
Ginsburg	13	48	7
Breyer	14	48	6
Roberts	5	58	3
Alito	3	27	3
O'Connor	0	18	0

As shown above, Chief Justice Roberts also wrote or joined the fewest separate opinions of any of the Justices sitting during the full 2005 Term.[19] Chief Justice Roberts signed concurring opinions in only three cases and dissented in only five more. This gives him a dissent rate of only 7.3%, the lowest rate of any full-term Justice. This rate also is significantly lower than Justice Rehnquist's dissent rate in his first term as Chief Justice: in 1986, Chief Justice Rehnquist dissented over 23% of the time. In fact, among the Justices sitting in the 2005 Term, Justice Rehnquist's 1986 record is closest to Justice Stevens's, who had a 2005 Term dissent rate of 25%.

Despite this relatively high level of agreement among the Justices, there were distinct areas of disagreement. A closer examination of the issue areas and ideological direction of the 2005 cases, listed by the vote margin by which the cases were decided, illustrates this point:

18. *League of United Latin American Citizens v. Perry*, 126 S.Ct. 2594 (2006), is excluded from this analysis. Because of the extremely fractured opinion, it was not possible to identify the Justices' opinions as either concurring, dissenting, or majority (most justices signed both concurring and dissenting opinions).

19. Justice Alito did not join the Court until January 31, 2006, and, thus, did not serve the entire 2005 Term. *See* U.S. Supreme Court webpage, *supra* note 16.

Table 8[20]
Vote Margins in Cases with Ideologically Conservative Outcomes

	Vote in the Case									Total
	5-2	5-4	6-2	6-3	7-1	7-2	8-0	8-1	9-0	
Crim Pro	0	3	0	2	0	2	2	1	3	13
Civil Rights	0	0	1	1	0	1	1	1	0	5
First Amend.	1	1	1	1	0	0	1	0	0	5
Due Process	0	0	0	1	0	0	0	0	0	1
Attorneys	0	0	0	0	0	0	0	0	1	1
Econ Activity	0	1	0	0	1	2	1	0	2	7
Judicial Pwr	0	1	0	1	0	1	0	0	3	6
Misc	0	0	0	0	0	0	0	0	1	1
Total	1	6	2	6	1	6	5	2	10	39

Table 9
Vote Margins in Cases with Ideologically Liberal Outcomes

	Vote in the Case									Total
	5-3	5-4	6-2	6-3	7-1	7-2	8-0	8-1	9-0	
Crim Pro	1	1	0	1	0	0	1	0	3	7
Civil Rights	1	0	0	0	0	0	1	0	1	3
First Amend.	0	0	0	0	0	0	1	0	0	1
Due Process	1	0	0	0	0	0	0	0	0	1
Attorneys	0	0	0	1	0	0	0	0	0	1
Econ Activity	0	0	0	0	1	0	3	0	4	8
Judicial Pwr	0	1	0	0	0	0	0	0	2	3
Misc	0	1	0	0	0	0	0	0	2	3
Total	3	3	0	2	1	0	6	0	12	27

As shown above, the Justices disagreed most frequently in criminal procedure cases. Criminal procedure cases generating conservative outcomes were the most controversial, with five of these cases decided by either a five to four or a six to three vote margin. Criminal procedure cases yielding liberal results contributed an additional three narrowly decided cases. Conservative First Amendment cases also were quite contentious, with four of those cases decided by a six to three vote margin or closer (almost certainly not coincidentally, these two issue areas—criminal procedure and First Amendment, particularly

20. The three ideologically uncodable cases are excluded from this analysis. *See supra* note 12.

election law regulation—were two of the issue areas in which Justice O'Connor often cast a decisive vote).[21]

Liberal economic activity cases, in contrast, were the least contested, with all eight cases in that issue area decided by a seven to one or higher vote margin. The Court's federalism cases—an area that has generated numerous five to four decisions in recent years—also were surprisingly uncontroversial in the 2005 Term. Of the three federalism cases decided by the Court, two were unanimously decided in favor of the federal government.[22] A third, however, holding that state agencies do not enjoy sovereign immunity in certain bankruptcy proceedings was decided by a five to four vote with Justice O'Connor casting the decisive fifth vote against the State.[23]

This relative unity should not obscure a key point: buried within the list of five to three, five to four, and six to three decisions are many of the most important decisions of the 2005 Term. These decisions include: *Hamdan v. Rumsfeld*,[24] which limited the President's ability to unilaterally lower procedural protections available to individuals detained in the "war on terror;" *League of United Latin American Citizens v. Perry*,[25] which upheld in part and struck down in part the mid-decade Texas redistricting orchestrated by former House Majority Leader Tom Delay; *Rapanos v. United States*,[26] which involved the ability of the federal Environmental Protection Agency to regulate certain wetlands; and *Gonzales v. Oregon*,[27] which held that the federal Justice Department did not have statutory authority to block physician-assisted suicides. It is thus clear that despite Chief Justice Roberts' success in achieving a higher *overall* degree of unity than seen in the 2004 Term, he has been unable and perhaps unwilling to bring greater unity to the Court in precisely the high-profile cases where unity arguably would be the most beneficial.

C. "Important" Cases

My analysis to this point has not attempted to distinguish between the relative importance of the cases decided during the Court's 2005 Term. All

---

21. *See The Rehnquist Court*, *supra* note 3.

22. *See* N. Ins. Co. of N.Y. v. Chatham County, Ga., 126 S. Ct. 1689, 1695 (2006) (holding that a political subdivision that is not an "arm of the State" cannot claim Eleventh Amendment immunity); United States v. Georgia, 126 S. Ct. 877, 882 (2006) (holding that Congress has power under Section Five of the Fourteenth Amendment to abrogate state sovereign immunity for conduct that actually violates the Fourteenth Amendment).

23. Cent. Va. Cmty. Coll. v. Katz, 126 S. Ct. 990, 994 (2006) (holding, over the dissents of Chief Justice Roberts, Justices Thomas, Scalia, and Kennedy, that a bankruptcy trustee's proceeding to set aside the debtor's preferential transfers to state agencies was not barred by state sovereign immunity). Justice Alito had not yet replaced Justice O'Connor when the Court decided *Katz*; Justice O'Connor voted with the majority. *See* U.S. Supreme Court webpage, supra note 16.

24. 126 S. Ct. 2749, 2798 (2006).

25. 126 S. Ct. 2594, 2626 (2006).

26. 126 S. Ct. 2208, 2215, 2219 (2006).

27. 126 S. Ct. 904, 925 (2006).

Supreme Court decisions however are not created equal; some decisions clearly are much more significant than others.[28] In recognition of this, this section identifies the Term's most important cases and examines the Court's record in those particular cases.

I have identified sixteen "Important Cases" decided in the 2005 Term. I selected these cases through a combination of objective and subjective measures. First, I used two objective measures: the number of amicus curiae briefs filed in each case and the amount of media attention generated by each case.[29] Cases with relatively high numbers of either (or both) were deemed important. I then exercised some subjective editorial discretion, removing one case that the objective measures identified as important and adding three cases not captured by those measures. I removed from the list *Ayotte v. Planned Parenthood of Northern New England*,[30] a highly anticipated abortion case challenging the constitutionality of New Hampshire's parental consent law. *Ayotte* did not live up to expectations because the Court, in a unanimous opinion, chose not to address the constitutional claim and instead remanded the case on narrow, remedial grounds.[31] The three cases not captured by the objective measures, but whose substantive importance nonetheless warranted their inclusion as Important Cases, are the following: *Hudson v. Michigan*,[32] a five to four criminal procedure decision written by Justice Scalia holding that a violation of the "knock-and-announce" rule did not require the suppression of evidence found in the resulting search; *Central Virginia Community College v. Katz*,[33] a five to four federalism decision written by Justice Stevens holding that state agencies did not enjoy sovereign immunity from claims of preferential transfers made by a bankruptcy trustee in a Chapter 11 proceeding; and *Georgia v. Randolph*,[34] a five to three criminal procedure decision written by Justice Souter holding that consent to search a marital home was invalid when one spouse consented but the other spouse, also present on the premises, refused to consent.

This process resulted in the following list of sixteen Important Cases:

---

28. Various methods exist to measure the importance of particular Supreme Court opinions. For one interesting example, see Thomas A. Smith, *The Web of Law*, San Diego Legal Studies Research Paper No. 06-11 (Spring 2005), papers.ssrn.com/sol3/papers.cfm?abstract-id=642863 (using mathematical networking theories to illustrate how U.S. case law clusters over time around certain "hub" cases).

29. See Appendix B, *infra* p. 638. Amicus curiae briefs were counted using Westlaw's "Briefs and Other Related Documents" data. Media attention was measured by counting the number of times a case name (searched for as a phrase in most cases) was mentioned in Lexis Nexis's "News" database in the past two years.

30. 126 S. Ct. 961 (2006).

31. *See id.* at 969.

32. 126 S. Ct. 2159, 2168 (2006).

33. 126 S. Ct. 990, 994 (2006).

34. 126 S. Ct. 1515, 1519 (2006).

Table 10
Important Cases in 2005 Term

Case	Vote	Direction	Majority Writer	Issue Area	Non-Majority Opinions
DaimlerChrysler Corp. v. Cuno[35]	9-0	Conservative	Roberts	Judicial Power	1
S.D. Warren Co. v. Me. Bd. of Envtl. Prot.[36]	9-0	Liberal	Souter	Economic Activity	0
eBay Inc. v. MercExchange, L.L.C.[37]	9-0	n/a	Thomas	Economic Activity	2
Burlington N. Ry. v. White[38]	9-0	Liberal	Breyer	Civil Rights	1
Rapanos v. United States[39]	5-4	Conservative	Scalia	Economic Activity	4
Garcetti v. Ceballos[40]	5-4	Conservative	Kennedy	First Amend.	3
Rumsfeld v. Forum for Academic & Inst. Rights[41]	8-0	Conservative	Roberts	First Amend.	0
Arbaugh v. Y&H, Corp.[42]	8-0	Liberal	Ginsburg	Civil Rights	0
Gonzales v. Oregon[43]	6-3	Liberal	Kennedy	Privacy	2
Schaffer v. Weast[44]	6-2	Conservative	O'Connor	Civil Rights	3
League of United Latin Am. Citizens v. Perry[45]	5-4	n/a	Kennedy	Civil Rights	5
Randall v. Sorrell[46]	6-3	Conservative	Breyer	1st Am	5
Hamdan v. Rumsfeld[47]	5-3	Liberal	Stevens	Civil Rights	5
Hudson v. Michigan[48]	5-4	Conservative	Scalia	Criminal Procedure	2
Cent. Va. Cmty. Coll. v. Katz[49]	5-4	Liberal	Stevens	Federalism	1
Georgia v. Randolph[50]	5-3	Liberal	Souter	Criminal Procedure	5

35. 126 S. Ct. 1854 (2006).
36. 126 S. Ct. 1843 (2006).
37. 126 S. Ct. 1837 (2006).
38. 126 S. Ct. 2405 (2006).
39. 126 S. Ct. 2208 (2006).
40. 126 S. Ct. 1951 (2006).
41. 126 S. Ct. 1297 (2006).
42. 126 S. Ct. 1235 (2006).
43. 126 S. Ct. 904 (2006).
44. 126 S. Ct. 528 (2006).
45. 126 S. Ct. 2594 (2006).
46. 126 S. Ct. 2479 (2006).
47. 126 S. Ct. 2749 (2006).
48. 126 S. Ct. 2159 (2006).
49. 126 S. Ct. 990 (2006).
50. 126 S. Ct. 1515 (2006).

As indicated above, these sixteen cases differ in some interesting respects from the Court's record in the 2005 Term overall. Most notably, criminal procedure cases, which constituted approximately 29% of the Court's case load in the 2005 Term as a whole, are underrepresented in the Important Cases: only two (12.5%) of the Important Cases involved this issue area. Perhaps more significantly, the Important Cases were much more closely decided than were the 2005 cases as a whole:

Table 11
Comparison of Vote Margins in Important Cases & All Cases—2005

Vote Margin	2005 Important Cases	2005 All Cases
5-2	0%	1.5%
5-3	12.5	4.4
5-4	31.3	14.5
6-2	6.3	2.9
6-3	12.5	11.6
7-1	0	2.9
7-2	0	8.7
8-0	12.5	17.4
8-1	0	2.9
9-0	25	33.3

As noted above, the Court decided less than 35% of all the 2005 Term cases by vote margins of six to three or closer. Among the Important Cases, however, 62.6% were so decided. Moreover, the Court decided unanimously on only 25% of the Important Cases, as opposed to more than 33% of the 2005 cases taken as a whole.

This relative lack of unity in the Term's Important Cases also is evidenced by the number of separate opinions written by the Justices in these cases:

Table 12[51] Percentage of Cases in Which Each Justice Wrote or Joined a Non-Majority Opinion—2005		
	Important Cases	All 2005 Term Cases
Stevens	50.0 %	36.2 %
O'Connor[52]	0	0
Scalia	37.6	26.1
Kennedy	37.5	23.1
Souter	37.5	21.7
Thomas	31.3	26.4
Ginsburg	50.1	27.5
Breyer	44.6	29.0
Roberts	37.6	11.6
Alito[53]	18.8	8.6

As shown above, Justice Ginsburg wrote or joined the greatest number of non-majority opinions in Important Cases, while Justice Stevens held that position in the 2005 Term overall. Among the Justices who served the full Term, Justice Thomas wrote or joined the fewest non-majority opinions in Important Cases. In the Term taken as a whole, Chief Justice Roberts had written or joined the fewest number of these opinions.

The Important Cases also differed from the 2005 cases overall by generating more liberal results. Of the ideologically codable Important Cases, 50% yielded liberal results, and 50% yielded conservative results. Compare this to the 2005 Term's overall ideological breakdown: as noted above, 56.5% of the Court's 2005 decisions were conservative, while only 39% were liberal.[54]

The ideological balance struck by the Court in its Important Cases likely relates to another difference between the Important Cases and the 2005 decisions overall: Justice Kennedy wrote a disproportionately higher number of majority opinions in the Important Cases. In addition to writing a larger percentage of majority opinions in Important Cases than in the Term's cases

51. To maintain consistency within this comparison, *League of United Latin American Citizens v. Perry* is excluded from this analysis. *See supra* note 18.

52. Justice O'Connor did not serve the full Term. *See* U.S. Supreme Court webpage, *supra* note 16.

53. Justice Alito did not serve the full Term. *See id.*

54. *See* Table 2, *supra* p. 609.

overall, Justice Kennedy also wrote more majority opinions in the Important Cases than any other Justice:

Table 13 Justices Who Wrote Majority Opinions in Important Cases & All Cases—2005		
Majority Opinion Writer	**Important Cases**	**All Cases**
Stevens	12.5 %	10.1 %
O'Connor	6.3	4.3
Scalia	12.5	13.0
Kennedy	**18.8**	**11.6**
Souter	12.5	10.1
Thomas	6.3	11.6
Ginsburg	6.3	10.1
Breyer	12.5	10.1
Roberts	12.5	11.6
Alito	0	5.8

Justice Kennedy's relative dominance of the Term's Important Cases supports the argument that he is now the functional center of the Roberts Court.

II. CHIEF JUSTICE ROBERTS AND JUSTICE ALITO

A. Ideological direction

The 2005 Term is interesting not just for the decisions made by the Court as a whole, but because it provides the first opportunity to evaluate the jurisprudence of Chief Justice Roberts and Justice Alito. Because Justice Alito did not join the Court until January 2006, there were only thirty-three decisions issued in the 2005 Term in which all of the currently sitting Justices participated. Within this body of cases, Chief Justice Roberts and Justice Alito voted for almost twice as many conservative as liberal outcomes.

Table 14 2005 Voting Direction by Justice			
	Conservative	**Liberal**	**Uncoded**
Scalia	22	9	2
Thomas	21	10	2
Roberts	**20**	**11**	**2**
Alito	**19**	**12**	**2**
Kennedy	19	12	2
Ginsburg	13	18	2
Breyer	11	20	2
Souter	10	21	2
Stevens	10	21	2

As shown above, both Roberts and Alito cast conservative-leaning votes in approximately 60% of these cases (60.6% and 57.5%, respectively). Notably, however, although both new Justices cast more conservative votes than the more liberal members of the Court (Justices Stevens and Breyer cast only ten and eleven conservative votes respectively), they did not cast as many conservative votes as Justices Scalia and Thomas: Justice Scalia cast the most conservative votes (twenty-two votes, or 66.6%), followed by Justice Thomas (twenty-one votes, or 63.6%). Interestingly, Justice Alito cast exactly the same number of liberal and conservative votes as did centrist Justice Kennedy.[55]

The issue areas in which Chief Justice Roberts and Justice Alito cast these votes are as follows:

Table 15 Ideological Votes by Issue Area—Justice Roberts (Alito)				
	Conservative	**Liberal**	**Uncoded**	**Total**
Crim. Pro.	8 (8)	5 (5)	0	13
Civil Rights	2 (2)	1 (1)	1	4
First Amend.	2 (2)	0 (0)	0	2
Due Proc.	1 (1)	0 (0)	0	1
Econ. Act.	3 (3)	2 (2)	1	6
Judicial Power	4 (3)	2 (3)	0	6
Federalism	0 (0)	1 (1)	0	1
Total	20 (19)	11 (12)	2	33

55. These votes were, however, not always cast in the same cases. *See* Appendix A, *infra* p. 630

As shown above, both of the new Justices cast a plurality of their conservative votes in criminal procedure cases (although, because of the quantity of cases heard in this area, it also is the largest single issue area in which both of the Justices cast liberal votes).

B. Judicial Alignment Rates

Chief Justice Roberts and Justice Alito had the highest judicial alignment rate of any of the justices. The sole case in which the two Justices reached opposing outcomes was *Empire Healthchoice Assurance, Inc. v. McVeigh,*[56] a judicial power case in which a majority comprised of Justices Stevens, Scalia, Thomas, Ginsburg, and Chief Justice Roberts held that the case involving a complex health insurance claim did not present a federal question. Justice Alito joined Justices Kennedy, Souter, and Breyer and voted in favor of the exercise of federal jurisdiction.

Despite this high rate of agreement between Chief Justice Roberts and Justice Alito, differences do exist between their ideological positions and also between their positions and the positions of the other Justices. These differences can be best examined by looking beyond majority/minority voting blocs and examining the rates at which the Justices joined each other's concurring and dissenting opinions.

Again examining only the cases in which all of the currently sitting nine Justices participated, the Justices' alignment rates are below. In calculating alignment rates, I opted against dividing the Justices into simple majority and dissenting blocs. Under my methodology, each Justice is counted as aligned with only one opinion per case, but that opinion can be a majority, concurrence, or dissent. For example, if Chief Justice Roberts, Justices Kennedy, Alito, Scalia and Thomas each join a majority opinion written by Chief Justice Roberts, but Justice Scalia writes a concurring opinion joined only by Justice Thomas, then Justices Scalia and Thomas will be counted as aligned with each other. However, neither will be counted as aligned with Justice Roberts.[57] If a Justice authors an opinion, then he or she always is aligned with that opinion.[58]

---

56. 126 S. Ct. 2121, 2127 (2006).

57. In cases in which a Justice joined more than one non-majority opinion, I made a subjective assessment about which secondary opinion was substantively more significant and aligned the Justice with that opinion. The only case of the thirty-three coded for this analysis that presented a serious problem was *Day v. McDonough,* in which Justice Breyer joined the dissents of both Justice Stevens and Justice Scalia. 126 S. Ct. 1675, 1685 (2006). Justice Stevens' dissent did not address the merits of the majority's decision. *Id.* It disagreed with the decision to enter judgment in the case while a related case remained pending. *Id.* Therefore, I aligned Justice Breyer's vote in this case with Justice Scalia's more substantive dissent.

58. This method of determining alignment actually made some of the most fractured decisions the easiest to code. For example, in the Texas political gerrymandering case, *League of United Latin American Citizens v. Perry,* six Justices authored opinions, leaving only Justices Thomas, Alito, and Ginsburg to be aligned elsewhere. I ultimately opted to group the Justices in

Unlike tabulating agreement rates by simply dividing Justices into majority and dissenting blocs, this methodology tracks the nuanced differences of opinion between the Justices as evidenced in their concurring and dissenting opinions.[59] It is thus a far more subtle way of measuring the Justices' analytical preferences. Under this methodology, the lack of alignment between Justices in any particular case (for example, the lack of alignment between Justice Scalia and Chief Justice Roberts in the example given above) is not necessarily evidence of actual *disagreement.* It means only that there was an opinion written in the case that a given Justice agreed with *more* than the other opinions.

Using this methodology, the rates of alignment between the Justices who served the entire 2005 Term were as follows:[60]

Table 16
Alignment Rates (%) of Justices in 2005

	Scalia	Kennedy	Souter	Thomas	Ginsburg	Breyer	Roberts	Alito
Stevens	**33.0**	42.0	61.0	36.0	54.5	54.5	45.5	36.0
Scalia	--	51.5	45.5	76.0	42.0	45.5	67.0	64.0
Kennedy		--	54.5	61.0	48.5	48.5	70.0	70.0
Souter			--	45.5	70.0	73.0	58.0	58.0
Thomas				--	42.0	45.5	76.0	73.0
Ginsburg					--	48.5	58.0	45.5
Breyer						--	57.5	48.5
Roberts							--	**82.0**

That Justices Stevens and Scalia were aligned the least often is hardly surprising. Much more interesting is the extraordinarily high level of alignment between Chief Justice Roberts and Justice Alito. As noted above, there is only one case, *Empire Healthchoice Assurance, Inc. v. McVeigh*,[61] in which these two Justices voted for different outcomes. There were only five additional cases, however, in which Chief Justice Roberts and Justice Alito were not perfectly aligned.[62]

---

that case as follows: 1) Kennedy; 2) Stevens; 3) Souter and Ginsburg; 4) Breyer; 5) Roberts and Alito; and 6) Scalia and Thomas.

59. More nuanced levels of agreement can be extrapolated even from unanimous cases (if, for example, a Justice writes a concurring opinion). Unanimous cases, therefore, are not excluded from my analysis, as they often are from similar works.

60. Unless a percentage falls precisely between two numbers, percentages are rounded to the nearest whole number.

61. 126 S. Ct. 2121 (2006).

62. *See* Randall v. Sorrell, 126 S. Ct. 2479, 2500 (2006) (a conservative First Amendment case striking down Vermont's campaign finance regulations in which Justice Alito concurs with

There are several possible explanations for Chief Justice Roberts' and Justice Alito's alignment. For example, two Justices nominated in quick succession by the same President and confirmed by the same Senate may reasonably be assumed to have very similar judicial philospophies; two new Justices joining a well-established group of Justices may find a natural affiliation with each other regardless of their ideological preferences; or both Justices as new justices may have chosen to exercise a similarly cautious approach during their first months on the Court. However, this extremely high alignment rate may suggest that Chief Justice Roberts and Justice Alito have a common jurisprudential approach distinct from that of the Court's more established, conservative Justices. This possibility certainly warrants scrutiny in the coming years.

Also interesting is the surprisingly high rate of alignment Chief Justice Roberts shares with Justices Souter and Ginsburg. Chief Justice Roberts aligned himself with these two Justices 58% of the time. He aligned himself with Justice Breyer almost as often, 57.5% of the time. This makes Justice Roberts' alignment rate with Clinton-appointees Justices Ginsburg and Breyer higher than the alignment rate between those two Justices themselves: Justices Ginsburg and Breyer aligned with each other in only 48.5% of the cases examined here. While it obviously is too soon to make informed predictions on this point, these alignment rates raise the intriguing possibility that Chief Justice Roberts may be willing to cross ideological lines—at least in low profile cases—and work with the Court's more liberal Justices to create a "minimalist majority" that counters the more "maximalist" jurisprudence of Justices Scalia and Thomas.[63]

CONCLUSION

This "by the numbers" examination of the 2005 Term reveals several interesting patterns. First, while the Roberts Court issued many more conservative opinions than did the Rehnquist Court in the 2004 Term, it was the 2004 Term, not the 2005 Term, that was inconsistent with the Court's

---

a majority opinion that Chief Justice Roberts joins); Dixon v. United States, 126 S. Ct. 2437, 2449 (2006) (a conservative criminal procedure case in which Justice Alito concurs with the majority opinion that Chief Justice Roberts joins); Burlington N. Santa Fe Ry. Co. v. White, 126 S. Ct. 2405, 2418 (2006) (a liberal civil rights case involving employment discrimination in which Justice Alito concurs with the majority opinion that Chief Justice Roberts joins); Rapanos v. United States, 126 S. Ct. 2208, 2235 (2006) (a conservative economic activity case involving regulation of certain wetlands by the federal Environmental Protection Agency in which Chief Justice Roberts concurs with the majority opinion that Justice Alito joins); eBay Inc. v. MercExchange, L.L.C., 126 S. Ct. 1837, 1841 (2006) (an ideologically uncoded economic activity case involving a patent dispute in which Chief Justice Roberts concurs with the majority opinion that Justice Alito joins).

63. For a discussion of judicial minimalism, see CASS R. SUNSTEIN, ONE CASE AT A TIME: JUDICIAL MINIMALISM ON THE SUPREME COURT (1999).

recent ideological record. Thus, it appears that the addition of (presumably) conservative Chief Justice Roberts and Justice Alito has not—or not yet—dramatically altered the ideological direction of the Court. Moreover, the most important cases yielded a higher percentage of liberal results than did the Term's cases overall. This may provide some reassurance to those who feared that Chief Justice Roberts and Justice Alito would move the Court sharply rightward.

Second, it appears that Chief Justice Roberts is making progress toward accomplishing his stated goal of bringing more unity to the Court. The Roberts Court issued fewer closely divided decisions in the 2005 Term than the Rehnquist Court had in 2004. Although this unity broke down in the Court's more important cases, Chief Justice Roberts himself contributed to the relative harmony by writing notably fewer concurring or dissenting opinions than the other Justices. Moreover, Chief Justice Roberts' apparent reluctance to write concurring or dissenting opinions, and the relatively high rates of alignment between himself and the Court's more liberal Justices, may indicate the new Chief Justice's willingness to sign on to the type of narrow, minimalist opinions that can garner broader support on the Court.

Finally, the extremely high alignment rate of the Court's two newest Justices shows that these Justices may have more in common with each other than with the other conservatives on the Court, an intriguing possibility that warrants ongoing attention as these Justices continue to serve together in coming years.

Appendix A

Citation	Vote	Direction	Value	Stevens	Scalia
126 S.Ct. 1675	5-4	liberal	judicial power	dissent	dissent
126 S.Ct. 1689	9-0	liberal	federalism	majority	majority
126 S.Ct. 1752	9-0	liberal	economic activity	majority	majority
126 S.Ct. 1735	9-0	liberal	judicial power	concur	majority
126 S.Ct. 1727	9-0	liberal	criminal procedure	majority	majority
126 S.Ct. 1854	9-0	conservative	judicial power	majority	majority
126 S.Ct. 1869	9-0	conservative	economic activity	majority	majority
126 S.Ct. 1843	9-0	liberal	economic activity	majority	majority
126 S.Ct. 1837	9-0	n/a	economic activity	concur	concur
126 S.Ct. 1943	9-0	conservative	criminal procedure	concur	majority
126 S.Ct. 2378	6-3	conservative	judicial power	dissent	majority
126 S.Ct. 2405	9-0	liberal	civil rights	majority	majority
126 S.Ct. 2422	8-1	conservative	civil rights	dissent	majority
126 S.Ct. 2437	7-2	conservative	criminal procedure	majority	concur
126 S.Ct. 2188	6-3	liberal	criminal procedure	majority	dissent
126 S.Ct. 2193	6-3	conservative	criminal procedure	dissent	majority
126 S.Ct. 2266	8-1	conservative	criminal procedure	majority	majority
126 S.Ct. 2208	5-4	conservative	economic activity	dissent	majority

Appendix A
(Continued)

Kennedy	Souter	Thomas	Ginsburg	Breyer	Roberts	Alito	OC
majority	majority	dissent	majority	dissent	majority	majority	.
majority	majority	majority	majority	majority	majority	majority	.
majority	majority	majority	majority	majority	majority	majority	.
majority	majority	majority	majority	majority	majority	majority	.
majority	majority	majority	majority	majority	majority	majority	.
majority	majority	majority	concur	majority	majority	majority	.
majority	majority	majority	majority	majority	majority	majority	.
majority	majority	majority	majority	majority	majority	majority	.
concur	concur	majority	concur	concur	concur	majority	.
majority	majority	majority	majority	majority	majority	majority	.
majority	dissent	majority	dissent	concur	majority	majority	.
majority	majority	majority	majority	majority	majority	concur	.
majority	majority	majority	majority	majority	majority	majority	.
concur	dissent	majority	majority	dissent	majority	concur	.
dissent	majority	dissent	majority	majority	majority	majority	.
majority	dissent	majority	majority	dissent	majority	majority	.
majority	majority	dissent	majority	majority	majority	majority	.
concur	dissent	majority	dissent	dissent	concur	majority	.

Appendix A (Continued)

Citation	Vote	Direction	Value	Stevens	Scalia
126 S.Ct. 2121	5-4	conservative	judicial power	majority	majority
126 S.Ct. 2557	5-4	liberal	criminal procedure	majority	majority
126 S.Ct. 2516	5-4	conservative	criminal procedure	dissent	concur
126 S.Ct. 2145	9-0	conservative	judicial power	majority	concur
126 S.Ct. 2159	5-4	conservative	criminal procedure	dissent	majority
126 S.Ct. 2096	9-0	liberal	criminal procedure	majority	majority
126 S.Ct. 1976	9-0	liberal	criminal procedure	majority	concur
126 S.Ct. 1991	7-2	conservative	economic activity	majority	concur
126 S.Ct. 1951	5-4	conservative	First Amend.	dissent	majority
126 S.Ct. 1515	5-3	liberal	criminal procedure	concur	dissent
126 S.Ct. 1494	8-0	conservative	criminal procedure	concur	majority
126 S.Ct. 1503	8-0	liberal	economic activity	majority	majority
126 S.Ct. 1297	8-0	conservative	First Amend.	majority	majority
126 S.Ct. 1281	8-0	liberal	economic activity	majority	majority
126 S.Ct. 1264	8-0	liberal	economic activity	majority	majority
126 S.Ct. 1276	8-0	conservative	economic activity	majority	majority
126 S.Ct. 1226	8-0	conservative	criminal procedure	majority	concur
126 S.Ct. 1235	8-0	liberal	civil rights	majority	majority

Appendix A
(Continued)

Kennedy	Souter	Thomas	Ginsburg	Breyer	Roberts	Alito	OC
dissent	dissent	majority	majority	dissent	majority	dissent	.
dissent	majority	dissent	majority	majority	dissent	dissent	.
majority	dissent	majority	dissent	dissent	majority	majority	.
majority	majority	majority	majority	majority	majority	majority	.
concur	dissent	majority	dissent	dissent	majority	majority	.
majority	majority	majority	majority	majority	majority	majority	.
majority	majority	majority	majority	majority	majority	majority	.
majority	majority	dissent	majority	dissent	majority	majority	.
majority	dissent	majority	dissent	dissent	majority	majority	.
majority	majority	dissent	majority	concur	dissent	.	.
majority	concur	majority	concur	majority	majority	.	.
majority	majority	majority	majority	majority	majority	.	.
majority	majority	majority	majority	majority	majority	.	.
majority	majority	majority	majority	majority	majority	.	.
majority	majority	majority	majority	majority	majority	.	.
majority	majority	majority	majority	majority	majority	.	.
majority	majority	concur	majority	majority	majority	.	.
majority	majority	majority	majority	majority	majority	.	.

Appendix A (Continued)

Citation	Vote	Direction	Value	Stevens	Scalia
126 S.Ct. 1252	7-1	liberal	economic activity	majority	majority
126 S.Ct. 1246	8-0	conservative	civil rights	majority	majority
126 S.Ct. 1204	7-1	conservative	economic activity	majority	majority
126 S.Ct. 1211	8-0	liberal	First Amend.	majority	majority
126 S.Ct. 980	7-2	conservative	judicial power	dissent	majority
126 S.Ct 990	5-4	liberal	federalism	majority	dissent
126 S.Ct. 952	9-0	conservative	judicial power	majority	majority
126 S.Ct. 969	9-0	conservative	criminal procedure	majority	majority
126 S.Ct. 961	9-0	conservative	misc	majority	majority
126 S.Ct. 941	8-0	n/a	misc	majority	majority
126 S.Ct. 904	6-3	liberal	privacy	majority	dissent
126 S.Ct. 884	5-4	conservative	criminal procedure	dissent	majority
126 S.Ct. 846	9-0	conservative	criminal procedure	concur	majority
126 S.Ct 860	7-2	conservative	economic activity	dissent	majority
126 S.Ct. 877	9-0	liberal	federalism	concur	majority
126 S.Ct. 699	9-0	conservative	economic activity	majority	concur
126 S.Ct. 704	9-0	conservative	attorneys	majority	majority
126 S.Ct. 676	7-2	conservative	civil rights	majority	majority

Appendix A
(Continued)

Kennedy	Souter	Thomas	Ginsburg	Breyer	Roberts	Alito	OC
majority	majority	dissent	majority	majority	majority	.	.
majority	majority	majority	majority	majority	majority	.	.
majority	majority	dissent	majority	majority	majority	.	.
majority	majority	majority	majority	majority	majority	.	.
dissent	majority	majority	majority	majority	majority	.	majority
dissent	majority	dissent	majority	majority	dissent	.	majority
majority	majority	majority	majority	majority	majority	.	majority
majority	concur	majority	majority	concur	majority	.	majority
majority	majority	majority	majority	majority	majority	.	majority
majority	majority	.	majority	majority	majority	.	majority
majority	majority	dissent	majority	majority	dissent	.	majority
majority	dissent	majority	dissent	dissent	majority	.	majority
majority	majority	majority	majority	majority	majority	.	majority
majority	majority	dissent	majority	majority	majority	.	majority
majority	majority	majority	concur	majority	majority	.	majority
majority	majority	majority	majority	majority	majority	.	majority
majority	majority	majority	majority	majority	majority	.	majority
dissent	majority	majority	dissent	majority	majority	.	majority

Appendix A (Continued)

Citation	Vote	Direction	Value	Stevens	Scalia
126 S.Ct. 606	9-0	liberal	judicial power	majority	majority
126 S.Ct. 528	6-2	conservative	civil rights	concur	majority
126 S.Ct. 510	9-0	liberal	economic activity	majority	majority
126 S.Ct. 514	9-0	liberal	economic activity	majority	majority
126 S.Ct. 1695	5-2	conservative	First Amend.	majority	majority
126 S.Ct. 2064	8-0	liberal	criminal procedure	majority	concur
126 S.Ct. 1708	5-3	liberal	due process	majority	dissent
126 S.Ct. 2572	6-2	conservative	First Amend.	dissent	concur
126 S.Ct. 2594	5-4	n/a	civil rights	split	split
126 S.Ct. 2669	6-3	conservative	criminal procedure	dissent	majority
126 S.Ct. 2455	6-3	conservative	civil rights	dissent	majority
126 S.Ct. 2479	6-3	conservative	First Amend.	dissent	concur
126 S.Ct. 2546	7-2	conservative	criminal procedure	dissent	majority
126 S.Ct. 2709	6-3	conservative	due process	dissent	majority
126 S.Ct. 2749	5-3	liberal	civil rights	majority	dissent
Total Cases	69	69	69	69	69

Appendix A
(Continued)

Kennedy	Souter	Thomas	Ginsburg	Breyer	Roberts	Alito	OC
majority	majority	majority	majority	majority	majority	.	majority
majority	majority	majority	dissent	dissent	.	.	majority
majority	majority	majority	majority	majority	majority	.	majority
majority	majority	majority	majority	majority	majority	.	majority
majority	majority	majority	dissent	dissent	.	.	.
majority	majority	concur	majority	majority	concur	.	.
dissent	majority	dissent	majority	majority	majority	.	.
majority	majority	concur	dissent	majority	majority	.	.
majority	split	split	split	split	split	split	.
majority	dissent	majority	concur	dissent	majority	majority	.
majority	dissent	majority	concur	dissent	majority	majority	.
concur	dissent	concur	dissent	majority	majority	concur	.
concur	majority	majority	dissent	majority	majority	majority	.
dissent	majority	majority	dissent	concur	majority	majority	.
concur	concur	dissent	concur	concur	-	dissent	.
69	69	68	69	69	67	34	18

Appendix B

Citation	Case Name	Media Mentions	Amici Briefs
126 S.Ct. 1675	Day v. McDonough	7	7
126 S.Ct. 2557	US v. Gonzalez-Lopez	11	1
126 S.Ct. 884	Brown v. Sanders	13	1
126 S.Ct. 2121	Empire Healthchoice v. McVeigh	17	3
126 S.Ct. 2594	League of United Latin American Citizens v. Perry	29	21
126 S.Ct 990	Central VA Community College v. Katz	33	7
126 S.Ct. 2516	Kansas v. Marsh	38	4
126 S.Ct. 1515	Georgia v. Randolph	60	3
126 S.Ct. 2159	Hudson v. Michigan	69	3
126 S.Ct. 1951	Garcetti v. Ceballos	91	12
126 S.Ct. 2208	Rapanos v. US	128	41
126 S.Ct. 2188	Youngblood v. WV	0	0
126 S.Ct. 2378	Woodford v. Ngo	4	7
126 S.Ct. 2193	Samson v.CA	7	9
126 S.Ct. 2709	Clark v. Arizona	16	6
126 S.Ct. 2455	Arlington Central School Dist. v. Murphy	26	5
126 S.Ct. 1708	Jones v. Flowers	28	1
126 S.Ct. 2669	Sanchez-Llamas v. OR	43	18
126 S.Ct. 2479	Randall v. Sorrell	63	23
126 S.Ct. 904	Gonzales v. OR	225	27
126 S.Ct. 2749	Hamdan v. Rumsfeld	547	55
126 S.Ct. 2546	Washington v. Recuenco	6	4
126 S.Ct. 2437	Dixon v. US	6	1
126 S.Ct. 1695	Hartman v. Moore	9	1
126 S.Ct. 980	Unitherm Food v. Swift-Eckrich	17	1
126 S.Ct. 676	Wagnon v. Prairie Band Potamwatomi Nation	23	11
126 S.Ct. 2572	Beard v. Banks	23	6
126 S.Ct 860	Volvo Trucks v. Reeder-Simco GMC	32	7

Citation	Case Name	Media Mentions	Amici Briefs
126 S.Ct. 1991	Anza v. Ideal Steel Supply	38	2
126 S.Ct. 528	Schaffer v. Weast	86	10
126 S.Ct. 1252	Dolan v. USPS	6	1
126 S.Ct. 1204	Buckeye Check Cashing v. Cardegna	8	12
126 S.Ct. 2422	Fernandez-Vargas v. Gonzales	13	1
126 S.Ct. 2266	Davis v. Washington	24	13
126 S.Ct. 510	US v. Olson	3	0
126 S.Ct. 969	Rice v. Collins	4	2
126 S.Ct. 952	Will v. Hallock	5	0
126 S.Ct. 1976	Zedner v. US	5	0
126 S.Ct. 846	Evans v. Chavis	6	1
126 S.Ct. 1246	Domino's Pizza v. McDonald	8	3
126 S.Ct. 1494	U.S. v. Grubbs	12	2
126 S.Ct. 1727	Holmes v.South Carolina	13	6
126 S.Ct. 699	Lockhart v. US	14	1
126 S.Ct. 606	Lincoln Property v. Roche	14	3
126 S.Ct. 704	Martin v. Franklin Capital Corp.	15	2
126 S.Ct. 1689	Northern Ins. v. Chatham Co	15	2
126 S.Ct. 514	IBP v. Alvarez	17	6
126 S.Ct. 1854	Daimler Chrysler v. Cuno	18	35
126 S.Ct. 1226	Oregon v. Guzek	18	2
126 S.Ct. 1943	Brigham City v. Stuart	24	7
126 S.Ct. 1264	Scheidler v. NOW	26	19
126 S.Ct. 2096	Hill v. McDonough	32	9
126 S.Ct. 1752	Ar Dept. of Health v. Ahlborn	32	3
126 S.Ct. 941	Wachovia v. Schmidt	33	4

Citation	Case Name	Media Mentions	Amici Briefs
126 S.Ct. 2145	Kircher v. Putnam Funds Trust	35	5
126 S.Ct. 877	US v. Georgia	38	9
126 S.Ct. 1869	Sereboff v. MidAtlantic Medical Services	38	10
126 S.Ct. 1503	Merrill Lynch v. Dabit	45	12
126 S.Ct. 1297	Rumsfeld v. Forum for Academic Rights	46	29
126 S.Ct. 1211	Gonzales v. O'Centro Espirita	47	10
126 S.Ct. 1276	Texaco v. Dahger	50	16
126 S.Ct. 1735	Marshall v. Marshall	55	10
126 S.Ct. 2064	House v. Bell	55	4
126 S.Ct. 1281	IL Tool Works v. Independent Ink	69	21
126 S.Ct. 1843	S.D. Warren v. Maine Enviro Prot	79	17
126 S.Ct. 1837	eBay v. MercExchange	84	36
126 S.Ct. 1235	Arbaugh v. Y&H Co.	89	4
126 S.Ct. 2405	Burlington No. v. White	91	11
126 S.Ct. 961	Ayotte v. Planned Parenthood	213	34

BETTER OFF UNBORN? AN ANALYSIS OF WRONGFUL BIRTH AND WRONGFUL LIFE CLAIMS UNDER THE AMERICANS WITH DISABILITIES ACT

DARPANA M. SHETH*

* Law clerk to the Honorable Jerome A. Holmes, United States Court of Appeals for the Tenth Circuit, 2006-2007. The views expressed in this Article are my own and do not necessarily reflect those of my employer, the government, or anyone acknowledged in this footnote. I would like to thank Robert D. Dinerstein at the Washington College of Law for his inspiration and advice while I was studying and writing in the area of disability law. This Article also benefited in countless ways from the helpful comments and thoughtful criticism of Manisha M. Sheth, Gregg T. Nunziata, and Jeremiah E. Goulka. I would like to dedicate this Article to my parents on their 35th anniversary for their loving support, enduring patience and endless inspiration.

INTRODUCTION

A recent study shows that more than 80 percent of babies prenatally diagnosed with Down syndrome are aborted.[1] In an age of increasing reliance on prenatal and genetic testing, should state tort law encourage reproductive choices that discriminate against offspring with actual or potential disabilities? Proponents of wrongful birth and wrongful life claims answer this question in the affirmative by supporting claims that allow monetary damages for the negligent deprivation of the choice to abort or not conceive a child with an actual or potential congenital disability. But the notion that one should be compensated under state tort law for the deprivation of the opportunity to prevent the birth of a child with disabilities is inherently incompatible with the ideals embodied in our nation's commitment to end discrimination against individuals with disabilities.

This Article argues that the genetic torts of wrongful birth and wrongful life violate the prohibition of discrimination against individuals with disabilities by public entities contained in Title II of the Americans with Disabilities Act. Although there has been some law review commentary on the torts of wrongful birth and wrongful life, especially in the context of federal abortion law, this Article is among the first to examine the viability of these claims under the Americans with Disabilities Act.

On the fourteenth anniversary of the passage of the Americans with Disabilities Act of 1990 ("ADA"),[2] President George W. Bush declared that despite "the important progress the ADA has brought about for our citizens and our Nation . . . millions of Americans with disabilities continue to face both physical barriers and false perceptions."[3] It may come as no surprise that these "physical barriers and false perceptions" [4] are perhaps strongest in contexts involving "quality of life" determinations. The legal (and later political) battle over the withholding of sustenance from Terri Schiavo—whom the majority of examining medical professionals diagnosed as being in a "persistent vegetative state"[5] and whom disability rights advocates characterized as an individual with

---

1. *See* Marc Morano, *New Study Fuels Controversy Over Down Syndrome Abortions*, CYBERCAST NEWS SERVICE, Apr. 5, 2005, http://www.cnsnews.com/ViewCulture.asp?Page=%5CCulture%5Carchive%5C200504%5CCUL20050405a.html; George F. Will, Editorial, *Eugenics By Abortion: Is Perfection an Entitlement?* WASH. POST, Apr. 14, 2005, at A27, *available at* http://www.washingtonpost.com/wp-dyn/articles/A51671-2005Apr13.html.

2. Pub. L. No. 101-336, 104 Stat. 327 (codified as amended at 42 U.S.C. §§ 12101 to 12213 (2000)).

3. Proclamation No. 7804, 69 Fed. Reg. 46,051 (July 26, 2004), *available at* http://www.whitehouse.gov/news/releases/2004/07/20040726-5.html.

4. *Id.*

5. Eugene Robinson, Editorial, *Life and Death in Florida*, THE WASHINGTON POST, Mar. 23, 2005, at A-15.

disabilities[6]—is the most recent illustration of our nation's inability to reconcile our pursuit of the freedom of self-determination with our goal of equal and fair treatment of individuals with disabilities.[7] This tension figures prominently in end-of-life cases where an individual's right to die with dignity is balanced against the competing societal interest in preserving the sanctity of life. As demonstrated by the Schiavo controversy, reconciling these interests becomes increasingly complicated in cases of substituted judgment, where "quality of life" decisions are made on behalf of an incompetent individual.[8]

In contrast, in the context of pregnancy and birth, many legal scholars believe that the tension between the right of self-determination (which is often oversimplified by limiting it to an individual's reproductive freedom) and the societal interest in protecting the rights of children with disabilities is less prominent.[9] In large measure, this belief is based on the assumption that parents will protect the best interests of their child. But this assumption raises numerous questions: If parents believe that the best interests of their child dictate that, rather than suffer through a life with severe disabilities, the child should never be born, can this choice still be viewed as furthering the right of self-determination? If the disability at issue is not life-threatening, such as

---

6. Steven J. Taylor, *Betraying Terri Schiavo: Disability Is Not A Reason To Die*, POST-STANDARD (Syracuse, N.Y.), May 25, 2004, at A9.

7. Theresa Marie "Terri" Schiavo had collapsed in her home in 1990 and experienced respiratory and cardiac arrest, and ultimately became dependant on a gastric feeding tube. *See* Respondent Michael Schiavo's Opposition to Application for Injunction at 4-5, Schiavo *ex rel.* Schindler v. Schiavo, 544 U.S. 945 (2005) (No. 04A-825), *available at* http://news.findlaw.com/hdocs/docs/schiavo/32405acluopp.pdf. In 1998, her husband and guardian Michael Schiavo petitioned the Florida state courts to remove her feeding tube. *Id.* at 5. Her parents opposed this action, asserting that her brain damage was not as extensive as claimed. *Id.* By March of 2005, the legal and political battle included fourteen appeals and numerous motions, petitions and hearings in the Florida state courts; five suits in federal district court; state legislation struck down by the Supreme Court of Florida; a federal congressional subpoena; federal legislation, known as the Palm Sunday Compromise; and four denials of certiorari from the United States Supreme Court. *Id.* at 1-2. Despite this intervention, Terri Schiavo's feeding tube was removed on March 18, 2005. She died thirteen days later. *See Attorney: Terri's Husband Cradled Her*, CNN, Mar. 31, 2005, http://www.cnn.com/2005/US/03/31/schiavo.deathbed/.

8. On January 3, 2007 the parents of a nine-year-old girl with static encephalopathy, a disorder that leaves her unable to move and with the cognitive capacity of an infant, announced that they had been giving her hormones to stunt the growth of their daughter and subjected her to a hysterectomy for quality-of-life reasons. *See* http://ashleytreatment.spaces.live.com/blog/cns!E25811FD0AF7C45C!1837.entry?_c=BlogPart.

The resulting (very public) debate has ranged from support for the parents to accusations of eugenics and worse. *See* http://ashleytreatment.spaces.live.com/.

9. *See, e.g.*, Patricia M.A. Beaumont, *Wrongful Life and Wrongful Birth, in* CONTEMPORARY ISSUES IN LAW, MEDICINE, AND ETHICS 99, 112 (Sheila A.M. McLean ed., 1996) (arguing that the principle of self-determination relied on in end-of-life cases and medical treatment decisions of newborns with severe disabilities requires recognition of wrongful life actions).

hereditary deafness, should such reproductive choices that discriminate against offspring with potential disabilities be encouraged or even permitted? If parents are denied the opportunity to decide between life with disability and non-existence by a physician's negligence in performing genetic testing, should the parents be compensated?

This Article argues that the genetic torts of wrongful birth and wrongful life violate the ADA's prohibition of discrimination against individuals with disabilities by public entities and cannot be reconciled with this nation's goals in enacting and enforcing the ADA. Part I defines wrongful birth and wrongful life causes of action, describes the legal theories underpinning these claims, and traces the development of these torts to their current status. Part II discusses the ADA's history and mandate and demonstrates how wrongful birth and wrongful life causes of action violate the prohibition against disability-based discrimination contained in Title II of the ADA. Part III critically examines the underlying policy justifications for recognizing wrongful birth and wrongful life claims and proposes alternative mechanisms that are compatible with the mandate and broader goals of the ADA. In concluding that wrongful birth and wrongful life claims should no longer be recognized, the Article calls for reconciling "quality of life" jurisprudence so that it reflects our nation's longstanding commitment to ending discrimination against individuals with disabilities at every stage of life.

I. WRONGFUL BIRTH AND WRONGFUL LIFE CLAIMS

A. Description of Wrongful Birth and Wrongful Life Claims

Wrongful birth and wrongful life causes of action are a unique subset of medical malpractice claims arising from a defendant's negligent failure to inform potential parents of the risk that their offspring may suffer from a congenital defect.[10] Wrongful birth and wrongful life claims arise in two factual scenarios. In the first scenario, plaintiffs assert that the healthcare provider's failure to diagnose or disclose actual or potential birth defects during pregnancy deprived them of the opportunity to abort the genetically defective child. Because this scenario involves the politically and emotionally charged issue of eugenic abortion, most commentators focus on the first factual scenario when analyzing wrongful birth and wrongful life claims.[11] Indeed, wrongful birth and wrongful life claims have been defined as exclusively involving the

10. Alexander J. Macones, Jr. et al., *Obstetrics and Gynecology Malpractice*, *in* 3 MEDICAL MALPRACTICE 17G-1, 17G.04[4] (Matthew Bender & Co. 2006).

11. *See, e.g.*, Christine Intromasso, *Reproductive Self-Determination in the Third Circuit: The Statutory Proscription of Wrongful Birth and Wrongful Life Claims as an Unconstitutional Violation of* Planned Parenthood v. Casey's *Undue Burden Standard*, 24 WOMEN'S RTS. L. REP. 101, 102, 119-20 (2003) (discussing the implications of wrongful birth and wrongful life claims on a woman's right to seek an abortion).

deprivation of the opportunity to abort a pregnancy.[12] In addition, wrongful birth and wrongful life claims can arise in the context of negligent genetic counseling. This second factual scenario arises when a healthcare provider insufficiently advises the plaintiffs as to the risks of birth defects prior to conception.[13] In order to assert a claim for wrongful birth or wrongful life under either scenario, the child with the congenital defect must be born alive; otherwise, the action converts to a wrongful death suit.[14]

Although jurists and scholars do not always use the terms consistently, typically wrongful birth claims are brought by the parents or legal representatives of the children born with birth defects, while wrongful life claims are brought by the children in their own right.[15] At first blush, the distinction between wrongful birth and wrongful life claims simply appears to be the plaintiff's identity; however, the actual distinction is rooted in the significantly different legal theories underlying each claim.[16] For example, each claim stems from a distinct duty owed by the attending physician.[17] In wrongful birth claims, the physician has allegedly breached a duty owed to the mother.[18] In contrast, successful wrongful life claims require the existence of a duty owed by a physician to a child who was not born or even conceived at the time of the alleged tortious act.[19] Increasingly, courts are willing to find that

---

12. *See* Minn. Stat. 145.424 (2005) (prohibiting, as wrongful birth and wrongful life claims, causes of action based "on the claim that but for the negligent conduct of another, a child would have been aborted").

13. *See, e.g.*, Turpin v. Sortini, 643 P.2d 954, 965-66 (Cal. 1982) (allowing a child to recover special damages where the defendant medical provider failed to advise parents, prior to conception, of the risk that offspring would suffer from hereditary deafness, and thereby deprived them of the opportunity to choose not to conceive the child).

14. *See* Gregory G. Sarno, Annotation, *Tort Liability for Wrongfully Causing One to Be Born*, 83 A.L.R.3d 15, 19 (1978).

15. *See Turpin*, 643 P.2d at 957 n.4; Arche v. United Stated Dep't of Army, 798 P.2d 477, 479 (Kan. 1990); Greco v. United States, 893 P.2d 345, 347 (Nev. 1995).

16. Mark Strasser, *Wrongful Life, Wrongful Birth, Wrongful Death, and the Right to Refuse Treatment: Can Reasonable Jurisdictions Recognize All But One?*, 64 MO. L. REV. 29, 43 (1999).

17. As a species of negligence actions, wrongful birth and wrongful life claims require the plaintiff to demonstrate the essential elements of negligence: the defendant breached a duty owed to the plaintiff, which caused the alleged injury. BARRY R. FURROW ET AL., HEALTH LAW 781 (West Publishing Co. 1995); *see also* GEORGE P. SMITH, II, BIOETHICS AND THE LAW 150 (University Press of America 1993) ("An action for wrongful life breaks down into the elements common to all negligence actions.").

18. *See* Walker v. Mart, 790 P.2d 735, 739 (Ariz. 1990) ("A physician . . . has a duty to inform parents about fetal problems and risks."); Harbeson v. Parke-Davis, Inc., 656 P.2d 483, 488 (Wash. 1983) (noting that wrongful birth refers to "an action based on an alleged breach of the duty of a health care provider to impart information or perform medical procedures with due care").

19. *See* Lininger v. Eisenbaum, 764 P.2d 1202, 1209 (Colo. 1988) ("The underlying assumption of [the plaintiff's] claim for wrongful life is that a physician owes a duty to a child who might be born in the future . . . ."); Viccaro v. Milunsky, 551 N.E.2d 8, 12 (Mass. 1990).

physicians owe a duty to individuals not yet in existence; however, such recognition is not universal.[20]

While wrongful birth claims fit comfortably within established concepts of duty, both wrongful birth and wrongful life claims are considered to be the most controversial pregnancy-related torts because they significantly depart from traditional negligence claims in other key respects. First, wrongful birth and wrongful life claims broaden the traditional element of proximate cause.[21] Traditionally, tort liability under a theory of negligence requires not only cause-in-fact or actual cause, but proximate cause, as well.[22] Although plaintiffs in wrongful birth and wrongful life claims assert that but for the physician's conduct the child with disabilities would not have been conceived or brought to term, they do not (and cannot) maintain that the physician's negligent conduct was the proximate cause of the birth defect. Wrongful birth and wrongful life claims are premised on the theory that the physician's negligent conduct deprived the parents of an informed decision as to whether to conceive a child (in the case of genetic counseling) or abort the pregnancy (in the case of failure to diagnose or disclose). Therefore, the causal link between the physician's conduct and the resulting injury is more attenuated than that present in traditional negligence actions.[23]

Some scholars have questioned this proximate causation analysis, claiming that it "misunderstands the nature of the injury [which] is not the birth of the child with anomalies, but the loss of the parent's ability to make informed reproductive decisions about the birth of that child."[24] Instead, they view the problem as simply one of damages valuation.[25] This conception is belied by the fact that the compensation sought by plaintiffs is not limited to the mere deprivation of a reproductive choice, but also includes compensation for the "injury" of raising a child with disabilities, such as the readily quantifiable costs of medical care and specialized education. Therefore, courts often reject

---

20. *Compare* Renslow v. Mennonite Hosp., 367 N.E.2d 1250, 1255 (Ill. 1977) (allowing a child to maintain an action against hospital and physician as a result of a negligent transfusion, even though the transfusion occurred eight years prior to child's conception), Walker v. Rinck, 604 N.E.2d 591, 592 (Ind. 1992) (allowing child to bring a cause of action against a laboratory and physician whose allegedly negligent acts occurred prior to conception of the child), *and* Pitre v. Opelousas Gen. Hosp., 530 So.2d 1151, 1157 (La. 1988) (recognizing "a legal duty to a child not yet conceived but foreseeably harmed by the negligent delivery of health care services to the child's parents"), *with* Miller v. Duhart, 637 S.W.2d 183, 187 (Mo. Ct. App. 1982) (refusing to find that the physicians owed a duty to a child when both the duty and the alleged breach of that duty arose before the child's conception).

21. *See* Turpin v. Sortini, 643 P.2d 954, 961 (Cal. 1982); Ellis v. Sherman, 515 A.2d 1327, 1328 (Pa. 1986).

22. *See* Palsgraf v. Long Island R.R. Co., 162 N.E. 99, 101-02 (N.Y. 1928) (Cardozo, J.).

23 *See, e.g.*, Hester v. Dwivedi, 733 N.E.2d 1161, 1167 (Ohio 2000) ("The law of negligence does not hold a defendant liable for damages that the defendant did not cause. The Hesters concede that Alicia's spina bifida was not caused by anything the appellees did.").

24. FURROW ET AL., *supra* note 17, at 785.

25. *See id.* at 781.

wrongful birth and wrongful life claims not because damages are too speculative to calculate, but rather because such claims lack the requisite proximal link between the defendant's negligence and the costs incurred by the plaintiffs as a result of their child's disability.[26]

Wrongful birth and wrongful life claims also depart from traditional negligence claims because they involve two specific and unique comparisons. First, these claims involve a comparison between the value of a life challenged by disabilities with that of a life that never came into existence.[27] Such claims are distinct from prenatal injury cases in which the child would have been born healthy were it not for the defendant's negligent act committed in utero. In rejecting the argument that wrongful life claims fall within traditional prenatal injury cases and do not require any attempted evaluation of a right not to be born, the Supreme Court of California stated:

> Because nothing defendants could have done would have given plaintiff an unimpaired life, it appears inconsistent with basic tort principles [of compensation] to view the injury for which defendants are legally responsible solely by reference to plaintiff's present condition without taking into consideration the fact that if defendants had not been negligent she would not have been born at all.[28]

Second, wrongful birth and wrongful life claims necessitate a comparison between the value of a life with disability with that of a life without disability. Even the less traditional torts of wrongful conception[29] and wrongful pregnancy,[30] in which parents of a healthy child seek recovery for the costs of an unplanned pregnancy and delivery, do not require valuation of a life with disabilities. In these causes of action, because the unplanned child is born healthy, the pregnancy and the delivery are the asserted injuries. In contrast, wrongful birth and wrongful life claims assert that the injury is the resulting life itself, because having a child with a disability could have been "prevented"

---

26. *See, e.g.*, Azzolino v. Dingfelder, 337 S.E.2d 528, 536 (N.C. 1985), *cert. denied*, 479 U.S. 835 (1986) (rejecting the argument that wrongful birth claims may be viewed as "traditional" negligence claims, and noting the lack of proximate causation between the injury and a doctor's allegedly negligent actions).

27. A plaintiff cannot claim a wrongful birth or wrongful life cause of action if, despite the doctor's negligence, a healthy child is born. *See* Strasser, *supra* note 16, at 33.

28. Turpin v. Sortini, 643 P.2d 954, 961 (Cal. 1982).

29. In a wrongful conception action, parents sue a physician for a negligent sterilization or contraception procedure, which results in the birth of a healthy child. *See* Sarno, *supra* note 14, at 21-24. Wrongful conception claims have also been referred to as wrongful pregnancy or wrongful birth by some courts.

30. In a wrongful pregnancy action, also known as an action for wrongfully causing the birth of a healthy child, parents sue a physician for a negligently performed abortion procedure, which results in the birth of a healthy child. *See id.* at 78-79.

through informed reproductive choices.[31] According to one disability rights scholar:

> The rationales courts use both to award and to deny recovery for "normal" children stand in contrast to those articulated by jurisdictions recognizing wrongful birth or wrongful life in the context of a child born with a genetic defect. While courts give heavy emphasis to the inherent benefits of rearing a child in the former, many courts ignore these benefits in the latter.[32]

Thus, wrongful birth and wrongful life claims tend to discount, or even nullify, the value of life with a disability.[33]

B. Development of Wrongful Birth and Wrongful Life Torts

At common law, no cause of action existed for wrongful birth or wrongful life.[34] Initially, courts were reluctant to recognize wrongful birth and wrongful life claims because they deviated from traditional negligence principles.[35] In the 1967 case of *Gleitman v. Cosgrove*,[36] one of the first cases to address wrongful birth and wrongful life claims, the Supreme Court of New Jersey denied recovery to a child and his parents, as guardians ad litem, for physicians' negligent failure to inform the parents of the possibility that their child would suffer from birth defects.[37] The mother testified that when she informed the defendant doctors that she had contracted German measles during the first month of her pregnancy, both doctors reassured her that German measles would have no effect on her child.[38] In fact, her unborn child had at least a twenty percent chance of contracting a birth defect as a result of the virus.[39] With respect to the child's claim, the Supreme Court of New Jersey

---

31. Hickman v. Group Health Plan, Inc., 396 N.W.2d 10, 17 (Minn. 1986) (Simonett, J., concurring) ("In [wrongful conception and wrongful birth] cases, the plaintiff would not have had the child but for the negligence of the doctor, but the interests sought to be compensated are different. . . . In the one, it is the right not to have an *unplanned* child, while in the other it is the right not to have an *unwanted* child." (emphasis added)).

32. Wendy F. Hensel, *The Disabling Impact of Wrongful Birth and Wrongful Life Actions*, 40 HARV. C.R.-C.L. L. REV. 141, 154 (2005).

33. *See* Phillips v. United States, 508 F. Supp. 537, 543 (D. S.C. 1980); Berman v. Allan, 404 A.2d 8, 12-13 (N.J. 1979).

34. *See* Sarno, *supra* note 14, at 20-24, 36.

35. *See, e.g.*, Gleitman v. Cosgrove, 227 A.2d 689, 692-93 (N.J. 1967), *abrogated by* Berman v. Allan, 404 A.2d 8, 14-15 (N.J. 1979).

36. 227 A.2d 689 (N.J. 1967).

37. *Gleitman*, 227 A.2d at 689, 691, 693.

38. *Id.* at 690. There was some factual dispute concerning whether the defendant had warned her that there would be "a 20 per cent chance her baby would have some defect." *Id.* But because the case involved an appeal from a directed verdict, plaintiff's testimony was taken as true. *See id.* at 691.

39. The plaintiffs' medical expert, Dr. Louis Fraulo, testified that "women who have

refused to recognize a legally cognizable injury because of the impossibility of measuring "the difference between his life with defects against the utter void of nonexistence."[40] The court also refused to permit recovery on the parents' claims.[41] In addition to the difficulty of measuring damages, "substantial policy reasons prevent[ed] this Court from allowing tort damages for the denial of the opportunity to take an embryonic life."[42] The court cited to a contemporary statement from the *New York Times* concerning "the rights of the fetus" when it stated: "The right to life is inalienable in our society. A court cannot say what defects should prevent an embryo from being allowed life . . . ."[43] The court also referenced Jonathan Swift's satirical essay, "A Modest Proposal," by commenting:

> The sanctity of the single human life is the decisive factor in this suit in tort. Eugenic considerations are not controlling. We are not talking here about the breeding of prize cattle. It may have been easier for the mother and less expensive for the father to have terminated the life of their child while he was an embryo, but these alleged detriments cannot stand against the preciousness of the single human life to support a remedy in tort.[44]

Ultimately, the court in *Gleitman* refused to recognize a valid claim for wrongful birth or wrongful life because of the "countervailing public policy supporting the preciousness of human life."[45]

Six years later, the United States Supreme Court's landmark decision in *Roe v. Wade*[46] sparked a resurgence in wrongful birth and wrongful life claims. Some courts opined that the federal legalization of abortion in *Roe v. Wade*, and the Supreme Court's recognition of a fundamental privacy right in "a woman's decision whether or not to terminate her pregnancy,"[47] provided the necessary legal and policy foundations for a claim that the parents would have exercised their right to abort their unborn fetus had the physician detected potential congenital defects in time.[48]

In 1988, the Supreme Court of Colorado recognized a claim for wrongful birth in *Lininger v. Eisenbaum* by reasoning that it was "merely applying common law negligence principles and not 'creating' a new cause of action."[49]

German measles in the first trimester of their pregnancy will produce infants with birth defects in 20 to 50 per cent of the cases." *Id.* at 690.

40. *Id.* at 692.
41. *Id.* at 693.
42. *Id.*
43. *Id.*
44. *Id.*
45. *Id.*
46. 410 U.S. 113, 163-65 (1973).
47. *Id.* at 153.
48. *See* Robak v. United States, 658 F.2d 471, 476 (7th Cir. 1981); Phillips v. United States, 508 F. Supp. 544, 550-51 (D. S.C. 1981); Berman v. Allan, 404 A.2d 8, 14 (N.J. 1979).
49. 764 P.2d 1202, 1208 (Colo. 1988) (en banc).

In *Lininger*, the defendant physicians informed the parents that their first son's blindness was not hereditary.[50] Relying on this advice, the parents conceived a second child, and later discovered that he was also blind.[51] Subsequently, both children were diagnosed with Leber's congenital amaurosis, a type of blindness that is hereditary.[52] The Supreme Court of Colorado upheld the parents' claim, reasoning that the physicians' misdiagnosis and failure to inform the parents of the hereditary nature of their first child's blindness caused the injury[53] because the parents "would not have conceived a second child (or would have terminated the pregnancy) had they accurately been apprised of the possibility that a second child would suffer the same affliction."[54]

In *Lininger*, the court failed to adequately address the defendant physicians' policy arguments that recovery should not be permitted under wrongful birth causes of action. First, the defendants argued that wrongful birth claims should not be recognized because such claims are particularly susceptible to fraud, but the court rejected this argument as not supported by the evidence.[55] Second, the defendants argued that that recognizing wrongful birth claims would negatively impact the children involved in such suits.[56] The court "fail[ed] to see how the parents' recovery of extraordinary medical and educational expenses, so as to minimize the detrimental effect of the child's impairment, is outweighed by any speculation about stigma that he might suffer."[57] The court's analysis, however, overlooked the fact that such stigma is not speculative because, in order to demonstrate causation, the parents must testify in open court that they would not have had the child. Moreover, the court neglected to distinguish between the stigma present at the individual level with the stigma that is present in the disability community as a whole; the minimal gains an individual litigant makes often come at the expense of the entire disability community, which suffers from the stigma of government-sanctioned discrimination, as explained more fully in Part II of this Article.[58]

Currently, twenty-three jurisdictions recognize a cause of action for wrongful birth.[59] These jurisdictions frequently cite policy rationales that

---

50. *Id.* at 1203-04.

51. *Id.* at 1204.

52. *Id.*

53. *Id.* at 1208.

54. *Id.* at 1205.

55. *Id.* at 1208. ("The defendants have not convinced us that wrongful birth claims are particularly susceptible to being asserted fraudulently, and even were we so persuaded, the possibility of fraud has not been shown to be so great as to justify extinguishing meritorious claims.").

56. *Id.* at 1207-08.

57. *Id.* at 1207.

58. See *infra* Part II.B.2.

59. Only Maine recognizes wrongful birth claims through legislative enactment. *See* ME. REV. STAT. ANN. tit. 24, § 2931 (2005). The remaining jurisdictions recognize wrongful birth claims through judicial common law. *See* Keel v. Banach, 624 So.2d 1022, 1029 (Ala. 1993); Univ. of Arizona Health Scis. Ctr. v. Superior Court, 667 P.2d 1294, 1297, 1299 (Ariz. 1983);

include deterring negligence in prenatal testing, preserving personal autonomy in reproductive choices, and compensating parents for extraordinary expenses associated with their child's disability. While courts are increasingly recognizing wrongful birth claims, they are reluctant to permit recovery under a wrongful life cause of action; only three states currently recognize a cause of action for wrongful life.[60]

In 1984, confronting facts substantially similar to those in *Gleitman*, the Supreme Court of New Jersey recognized a wrongful life claim.[61] In *Procanik v. Cillo*, an infant born with multiple birth defects brought suit against the physicians who failed to diagnose his mother's German measles in the first trimester of her pregnancy.[62] The parents' claim for wrongful birth was time-barred by New Jersey's two-year statute of limitations.[63] Recognizing that in the absence of the defendants' negligence, the plaintiff would not have been born, the Supreme Court of New Jersey conceded the manifest problems with objectively measuring the value of non-existence.[64] Consequently, the court held that the infant plaintiff should be compensated for actual medical costs attributable to his birth defects, but not for general damages such as emotional distress.[65] The court rejected the "metaphysical considerations" it had previously espoused in *Gleitman*, stating that its "decision to allow the recovery of extraordinary medical expenses is not premised on the concept that non-life is preferable to an impaired life, but is predicated on the needs of the living."[66] Accordingly, the court sought "only to respond to the call of the living for help in bearing the burden of their affliction."[67] In addition, the court cited two

---

Turpin v. Sortini, 643 P.2d 954, 962, 966 (Cal. 1982); Lininger v. Eisenbaum, 764 P.2d 1202, 1207 (Colo. 1988); Haymon v. Wilkerson, 535 A.2d 880, 885-86 (D.C. 1987); Garrison v. Med. Ctr. of Delaware, Inc., 581 A.2d 288, 290 (Del. 1989); Kush v. Lloyd, 616 So.2d 415, 424 (Fla. 1992); Goldberg v. Ruskin, 471 N.E.2d 530, 536 (Ill. App. Ct. 1984); Arche v. U.S. Dep't of Army, 798 P.2d 477, 481 (Kan. 1990); Reed v. Campagnolo, 630 A.2d 1145, 1151 (Md. 1993); Viccaro v. Milunsky, 551 N.E.2d 8, 11-12 (Mass. 1990); Eisbrenner v. Stanley, 308 N.W.2d 209, 213 (Mich. Ct. App. 1981); Greco v. United States, 893 P.2d 345, 349-51 (Nev. 1995); Smith v. Cote, 513 A.2d 341, 347 (N.H. 1986); Schroeder v. Perkel, 432 A.2d 834, 839 (N.J. 1981); Becker v. Schwartz, 386 N.E.2d 807, 813 (N.Y. 1978); Flanagan v. Williams, 623 N.E.2d 185, 188-89 (Ohio Ct. App. 1993); Owens v. Foote, 773 S.W.2d 911, 913 (Tenn. 1989); Jacobs v. Theimer, 519 S.W.2d 846, 850 (Tex. 1975); Naccash v. Burger, 290 S.E.2d 825, 830 (Va. 1982); Harbeson v. Parke-Davis, Inc., 656 P.2d 483, 488, 496 (Wash. 1983); James G. v. Caserta, 332 S.E.2d 872, 877 (W. Va. 1985); Dumer v. St. Michael's Hosp., 233 N.W.2d 372, 377 (Wis. 1975).

60. *Turpin*, 643 P.2d at 966; Procanik v. Cillo, 478 A.2d 755, 757 (N.J. 1984); *Harbeson*, 656 P.2d at 495.

61. *Procanik*, 478 A.2d at 757.

62. *Id.*

63. *Id.* at 758.

64. *Id.* at 760.

65. *Id.* at 762.

66. *Id.* at 763.

67. *Id.*

jurisdictions that had also recognized the parents' right to recover actual medical costs in wrongful birth claims,[68] and stated that there was little rationale for fashioning a different rule in cases brought by the impaired child. Although the court acknowledged that its award of special damages for actual medical costs but denial of general damages was somewhat illogical, it stated:

> Law is more than an exercise in logic, and logical analysis, although essential to a system of ordered justice, should not become a [sic] instrument of injustice. Whatever logic inheres in permitting parents to recover for the cost of extraordinary medical care incurred by a birth-defective child, but in denying the child's own right to recover those expenses, must yield to the injustice of that result.[69]

Thus, the New Jersey Supreme Court clarified that its decision was grounded on principles of fairness. Nevertheless, as explained more fully later in this Article, the so-called "fairness" of awarding damages in such situations is undermined by the fact that wrongful life actions only award damages when the child disavows his own self-worth.[70] Children with disabilities who do not regret their existence, but instead choose to affirm the value of their lives by not bringing wrongful life claims, do not receive any court-awarded financial assistance.

In light of this fundamental *unfairness*, it is no surprise that nine states statutorily prohibit either wrongful birth or wrongful life actions on public policy grounds.[71] Some jurisdictions have upheld the constitutionality of such statutory prohibitions on the grounds that they do not violate the right to due process or equal protection of the law and do not infringe upon a woman's right to have an abortion.[72] But these statutes frequently differ in how they define wrongful birth and wrongful life claims. Consequently, certain variations of

---

68. *Id.* at 762 (citing Turpin v. Sortini, 643 P.2d 954, 965 (Cal. 1982) and Harbeson v. Parke-Davis, 656 P.2d 483 (Wash. 1983)).

69. *Id.*

70. *See infra* Part II.B.2.

71. IDAHO CODE ANN. § 5-334 (1986) (prohibiting wrongful birth and wrongful life claims); IND. CODE § 34-12-1-1 (1998) (barring wrongful life claims); MINN. STAT. § 145.424 (1993) (prohibiting wrongful birth and wrongful life claims); MO. REV. STAT. § 188.130 (1992) (prohibiting wrongful birth and wrongful life claims); N.C. GEN. STAT. § 14-45.1(e) (1993) (prohibiting wrongful birth and wrongful life claims); N.D. CENT. CODE § 32-03-43 (1993) (barring wrongful life claims); 42 PA. CONS. STAT. § 8305 (1993) (prohibiting wrongful birth and wrongful life claims); S.D. CODIFIED LAWS §§ 21-55-1 to -2 (1987) (prohibiting wrongful birth and wrongful life claims); UTAH CODE ANN. § 78-11-24 (1993) (prohibiting wrongful birth and wrongful life claims).

72. Hickman v. Group Health Plan, Inc., 396 N.W.2d 10, 11, 13 (Minn. 1986) (upholding a state statute prohibiting wrongful birth and wrongful life causes of action because the statute did not violate due process or equal protection of the law or impose a significant burden on a woman's right to have an abortion); Edmonds v. W. Pennsylvania Hosp. Radiology Assocs., 607 A.2d 1083, 1086, 1088 (Pa. 1992) (same).

wrongful birth or wrongful life claims may be permitted in states that ostensibly bar these causes of action. For example, Minnesota's statute bars wrongful birth and wrongful life claims where the plaintiff claims that "but for the negligent conduct of another, the [fetus] would have been *aborted*."[73] By its terms, this statute would not preclude wrongful birth or wrongful life claims in which the plaintiff alleges that but for the defendant's negligence, the child would not have been *conceived*.[74]

II. WRONGFUL BIRTH AND WRONGFUL LIFE CLAIMS AS A VIOLATION OF THE ADA

Given the relatively widespread acceptance of wrongful birth claims and the increasing recognition of wrongful life claims, our system of law must evaluate the viability of these relatively new causes of action as interpreted under federal disability discrimination laws.

A. The Americans with Disabilities Act: A National Mandate Against Disability Discrimination

After decades of efforts to pass a comprehensive federal disability discrimination law, Congress enacted the Americans with Disabilities Act ("ADA") in 1990. The primary purpose of this milestone piece of civil rights legislation was "to provide a clear and comprehensive national mandate for the elimination of discrimination against individuals with disabilities."[75]

Prior to enacting the ADA, Congress held thirteen hearings and created a special task force to gather evidence from each state of the pervasive discrimination that was occurring against individuals with disabilities.[76] Based on its lengthy and comprehensive investigation, Congress concluded that individuals with disabilities have been

> subjected to a history of purposeful unequal treatment, and relegated to a position of political powerlessness in our society, based on characteristics that are beyond the control of such individuals and resulting from stereotypic assumptions not truly indicative of the individual ability of such individuals to participate in, and contribute to, society.[77]

Moreover, for the first time, Congress expressly deemed segregation of individuals with disabilities as a form of discrimination.[78] Consequently,

---

73. MINN. STAT. § 145.424 (1993) (emphasis added).

74. Strasser, *supra* note 16, at 54.

75. 42 U.S.C. § 12101(b)(1) (2000).

76. Bd. of Trustees of the Univ. of Ala. v. Garrett, 531 U.S. 356, 377 & app. A (2001) (Breyer, J., dissenting).

77. 42 U.S.C. § 12101(a)(7).

78. *Id.* § 12101(a)(2).

Congress adopted a civil-rights model for addressing disability issues by using the Civil Rights Act of 1964 as a foundation for drafting and interpreting the ADA.[79]

Title II of the ADA broadened the scope of protection that Congress had previously afforded to individuals with disabilities in the Rehabilitation Act of 1973.[80] Section 504 of the Rehabilitation Act ("section 504") bars any program or activity that receives federal funding from discriminating on the basis of disability.[81] One of the purposes of Title II was to extend the non-discrimination principles espoused in section 504 to state and local public entities, regardless of whether they received federal funding.[82] Specifically, Title II provides that "no qualified individual with a disability shall, by reason of such disability, be excluded from participation in or be denied the benefits of the services, programs, or activities of a public entity, or be subjected to discrimination by any such entity."[83] Accordingly, Title II affords broader protection to individuals with disabilities than did its predecessor, section 504, because it prohibits discrimination by state and local agencies regardless of whether they receive federal funding and, consequently, expands the field of potential plaintiffs and defendants.[84]

---

79. S. COMM. ON LABOR AND HUMAN RESOURCES, THE AMERICANS WITH DISABILITIES ACT OF 1989, S. Rep. No. 101-116, at 43 (1989), *reprinted in* 1 LEGISLATIVE HISTORY OF PUBLIC LAW 101-336 THE AMERICANS WITH DISABILITIES ACT, at 141 (1990) (outlining which remedies and procedures under the Civil Rights Act of 1964 are also available pursuant to the ADA); H.R. Rep. No. 101-485, pt.2, at 82-83 (1990), *reprinted in* 1990 U.S.C.C.A.N. 303, 365.

80. 29 U.S.C. §§ 701 to 794. The Rehabilitation Act of 1973 established the Rehabilitation Services Administration, authorized vocational, employment, and rehabilitation programs for individuals with disabilities, and provided for the barrier-free reconstruction of public facilities. Francis M. Dougherty, Annotation, *Who is "Individual with Handicaps" Under Rehabilitation Act of 1973*, 97 A.L.R. FED. 40, 45-46 (2005).

81. Rehabilitation Act of 1973 § 504, 29 U.S.C. § 794(a) (2000).

82. Ann K. Wooster, Annotation, *When Does a Public Entity Discriminate Against Disabled Individuals in its Provision of Services, Programs, or Activities Under the Americans with Disabilities Act,* 163 A.L.R. FED. 339, 360 (2004).

83. 42 U.S.C. § 12132.

84. Schierberl v. Bird, 848 F. Supp. 303, 314 (D. N.H. 1994).

B. *Applying Title II of the ADA to Wrongful Birth and Wrongful Life Claims*[85]

Wrongful birth and wrongful life claims contain implicit characterizations of individuals with disabilities that violate both the spirit and the letter of the ADA. To illustrate, a claim for wrongful birth requires a state court to hold that a parent who has been denied the opportunity to abort a child with disabilities has suffered a compensable legal wrong. Likewise, by recognizing a wrongful life cause of action, a state court must determine that the infant with disabilities has suffered a compensable legal wrong by being born. Permitting recovery under such claims contradicts the clear mandate of Title II, which prohibits public entities from discriminating against individuals on the basis of their disability.

Consequently, wrongful birth and wrongful life claims violate Title II of the ADA in three respects. First, wrongful birth and wrongful life claims discriminate against individuals protected by the ADA, namely the children with disabilities at the center of these causes of action. Second, these causes of action discriminate against children on the basis of their disabilities by encouraging eugenic abortions and discounting the value of life with a disability. Third, judicial and legislative recognition and state enforcement of wrongful birth and wrongful life claims constitute discrimination by a public entity.

1. Children Born with Congenital Defects Are Protected by Title II of the ADA

Children born with congenital defects are classified as individuals with disabilities; as such, they are members of Title II's protected class. The ADA defines "disability" as: "(A) a physical or mental impairment that substantially limits one or more of the major life activities of such individual; (B) a record of such an impairment; or (C) being regarded as having such an impairment."[86] This definition is materially the same as the definition of "handicapped individual" under section 504 of the Rehabilitation Act.[87] Under this broad

---

85. Several academics have argued that wrongful birth and wrongful life claims are inconsistent with the policies embraced by anti-discrimination laws. *See e.g.*, Lori B. Andrews & Michelle Hibbert, *Courts and Wrongful Birth: Can Disability Itself be Viewed as a Legal Wrong?*, *in* AMERICANS WITH DISABILITIES: EXPLORING IMPLICATIONS OF THE LAW FOR INDIVIDUALS AND INSTITUTIONS 318, 318 (Leslie Pickering Francis & Anita Silvers eds. 2000); Hensel, *supra* note 32, at 141. But this Article is among the first to examine the legality of these actions under the ADA.

86. 42 U.S.C. § 12102(2).

87. *See* 29 U.S.C. § 705(9) (defining "disability" as "a physical or mental impairment that constitutes or results in a substantial impediment to employment" or an impairment "that substantially limits one or more major life activities"). In 1992, Congress amended the Rehabilitation Act to replace its use of the term "handicap" with the term "disability."

definition of disability, it would seem clear that newborn infants and children with congenital birth defects are protected by the ADA as individuals with disabilities. Nevertheless, there has been litigation in the context of section 504 of the Rehabilitation Act questioning whether children with congenital defects are "handicapped individuals," or individuals with disabilities, under the Act.[88]

In *United States v. University Hospital, State University of New York at Stony Brook* (hereinafter "*University Hospital*"), the Second Circuit held that an infant could be considered a "handicapped individual" within the meaning of the Rehabilitation Act.[89] In *University Hospital*, the United States Department of Health and Human Services, pursuant to its authority under section 504, sought to obtain the medical records of a severely deformed newborn infant to determine whether she was being discriminatorily denied medically indicated treatment on the basis of her disabilities.[90] The defendant hospital argued that, rather than her numerous physical and mental impairments, it was actually her status as an infant that substantially limited her major life activities.[91] The Second Circuit rejected this narrow reading and noted that there was no "explicit indication in the statute or regulations that 'major life activities' should be defined only with reference to adults."[92] At the least, the infant was *regarded* as having an impairment that substantially limits her major life activities under section 504's definition of "handicapped individual."[93] Ultimately, the court determined that the infant was not "otherwise qualified" as eligible for services because the infant's parents refused to consent to treatment.[94]

Under the ADA, courts have scrutinized the threshold determination of disability more closely.[95] Typically, children with minor birth defects do not

---

Rehabilitation Act Amendments of 1992, Pub. L. No. 102-569, 106 Stat. 4344 (1992). Before this revision, the Rehabilitation Act used the term "handicap" and the ADA used the term "disability," but the two terms were intended to be synonymous. *See* 29 U.S.C. § 706(8)(B)(i) (1988) (defining "individual with handicaps" as, *inter alia*, a person with "a physical or mental impairment which substantially limits one or more of such person's major life activities").

88. *See* Bowen v. Am. Hosp. Ass'n, 476 U.S. 610, 624 (1986) (stating in dicta that children with congenital defects are "handicapped individuals"); United States v. Univ. Hosp., State Univ. of N.Y. at Stony Brook, 729 F.2d 144, 155-56 (2d Cir. 1984).

89. *Univ. Hosp.*, 729 F.2d at 155.

90. *Id.* at 147-48.

91. *Id.* at 155.

92. *Id.*

93. *Id.*; *see also Bowen*, 476 U.S. at 624 (holding that the term "handicapped individual" as used in section 504 includes an infant who is born with a congenital defect).

94. *Univ. Hosp.*, 729 F.2d at 160; *see also infra* notes 96-97 and accompanying text (discussing the requirement that individuals be "qualified" for the allegedly discriminatory policy or service).

95. *See* NATIONAL COUNCIL ON DISABILITY, RIGHTING THE ADA 43 (2004), http://www.ncd.gov/newsroom/publications/2004/pdf/righting_ada.pdf (discussing the more restrictive interpretations courts have given the identical definition of disability under ADA than

qualify as proper subjects of wrongful birth or wrongful life claims because they have suffered only a *de minimis* injury; therefore, these children lack one of the essential elements of traditional negligence causes of action.[96] The majority of wrongful birth and wrongful life claims involve children with physical or mental impairments that substantially limit one or more major life activities, including cystic fibrosis,[97] Tay-Sachs disease,[98] and Down syndrome.[99] Thus, children at the center of wrongful birth and wrongful life claims who are afflicted with such severe diseases qualify as individuals with disabilities under the ADA.

There may be a group of children whose impairments are not so severe as to fall within the first prong of the disability definition, yet not so minor as to be precluded from recovering under a wrongful birth or wrongful life cause of action. Nevertheless, such individuals could still satisfy the third prong of the disability definition as individuals who are perceived to have a physical or mental impairment that substantially limits a major life activity. Such individuals would be regarded as having a disability not only by the parents who bring the wrongful birth claim but also by the courts adjudicating that claim.

A more complex issue involves interpretation of the term "qualified" under Title II and how it applies to individuals in the context of wrongful birth and wrongful life claims. Title II defines "qualified individual with a disability" as

> an individual with a disability who, with or without reasonable modifications to rules, policies, or practices, the removal of architectural, communication, or transportation barriers, or the provision of auxiliary aids and services, meets the essential eligibility requirements for the receipt of services or the participation in programs or activities provided by a public entity.[100]

Much of the litigation under Title II and its predecessor, section 504, involves challenges to allegedly discriminatory denials of employment, admission to educational programs, or access to public services. In such

---

under section 504).

96. *See supra* note 17.

97. *See* Schroeder v. Perkel, 432 A.2d 834, 842 (N.J. 1981) (holding that defendants could be held liable for incremental medical costs of two children born with cystic fibrosis where they negligently failed to diagnose the first child's condition in time to prevent second pregnancy or in time to abort the second child).

98. *See* Naccash v. Burger, 290 S.E.2d 825, 827 (Va. 1982) (recognizing a claim for wrongful birth where physicians negligently failed to discover that fetus had Tay-Sachs disease, "an invariably fatal disease of the brain and spinal cord that occurs in Jewish infants of eastern European ancestry").

99. *See* Haymon v. Wilkerson, 535 A.2d 880, 881 (D.C. 1987) (holding that a mother was entitled to recover extraordinary health care expenses attributable to her child's mental and physical abnormalities caused by Down syndrome).

100. 42 U.S.C. § 12131(2) (2000); *see also* 28 C.F.R. § 35.104 (2005) (restating this definition of a "qualified individual with a disability").

contexts, a plaintiff is a "qualified individual" if, with or without a reasonable accommodation, the plaintiff meets the minimum requirements of the program or service. For example, the court in *University Hospital* held that the infant with disabilities was not "otherwise qualified" for the medical services in question because her parents had not provided the requisite consent to the corrective surgery necessary to prolong her life.[101]

In contrast, this interpretation of a "qualified individual" does not readily transfer to the context of wrongful birth and wrongful life claims because there is no "program" or "service," per se. In an analogous situation involving disability discrimination by a healthcare provider, a Pennsylvania federal court stated that "simply because the functional standard adopted in employment and education does not neatly 'fit' in the medical benefits context does not necessarily mean the Rehabilitation Act does not apply to that context."[102] Similarly, though wrongful birth and wrongful life claims do not neatly fit into the ADA's eligibility standard, this aspect does not render the ADA inapplicable to the claims. Wrongful birth and wrongful life claims fall squarely within the last clause of section 12132 of Title II, which states that no qualified disabled person shall "be subjected to discrimination by any such entity."[103] This prong does not require the receipt of benefits or services or participation in programs in which eligibility must be considered. Insofar as there is an eligibility requirement to bring wrongful birth or wrongful life claims, disability itself is the minimum requirement to have standing to bring a cause of action for wrongful birth or wrongful life. By definition, these tort claims only apply in circumstances where an individual is born with a disability. Hence, Title II's protective ambit includes children with congenital defects and guards them from discrimination by public entities.

2. Wrongful Birth and Wrongful Life Claims Discriminate Against Children on the Basis of Their Disabilities

Wrongful birth and wrongful life claims discriminate on the basis of disability by encouraging eugenic abortions and discounting the value of a life with disability. Under the ADA, a showing of discriminatory *intent* is not necessary to sustain a violation. An "unjustifiable" disparate *impact* upon individuals with disabilities also can lead to a violation of the ADA.[104] Because

---

101. United States v. Univ. Hosp., State Univ. of N.Y. at Stony Brook, 729 F.2d 144, 156 (2d Cir. 1984). The court alternatively held that the infant with disabilities was not subjected to discrimination on the basis of her disability because the surgery was not performed due to the lack of parental consent, and not because of any discriminatory purpose. *Id.* at 160.

102. Woolfolk v. Duncan, 872 F. Supp. 1381, 1388 (E.D. Pa. 1995); *cf.* Bowen v. Am. Hosp. Ass'n, 476 U.S. 610, 624 (1986) (recognizing by plurality opinion that a hospital rule or state policy limiting a handicapped infant's "meaningful access" to medical services would be subject to challenge under section 504).

103. 42 U.S.C. § 12132.

104. *See* S. REP. NO. 101-116, at 44 (1989) (expressing intent that the ADA "be interpreted

wrongful birth and wrongful life claims are only available to children with congenital defects, these claims have a disparate impact on individuals with disabilities. Moreover, although wrongful birth and wrongful life claims provide some financial assistance to individual litigants, the overall impact of these claims is negative because they threaten individuals with disabilities, as well as the larger disability community.

Some jurists have recognized the way in which wrongful birth and wrongful life claims diminish the rights of individuals with disabilities. For example, Judge Wieand of the Superior Court of Pennsylvania observed that "our society has become increasingly aware and protective of the rights of the handicapped," and noted that "recogniz[ing] the validity of a wrongful life statute would be to make the handicapped a lower class of citizen."[105] Similarly, as Justice Winterscheimer of the Supreme Court of Kentucky astutely noted:

> A claim for wrongful life must be rejected because it would definitely discriminate against disabled persons and could lead to a eugenic culture where the "unfit" are made disposable. In effect, doctors and other healthcare professionals would be punished for not identifying and eliminating any disabled children in the womb. Such an approach is inherently dangerous. What "defect" will the law recognize as compensable? Who will draw the line as to what is severe and what is not severe? Will physical as well as mental impairments be involved?
>
> It is obvious that many of the so-called disabled can and do have lives of immense value to themselves and to others. Such a concept is clearly envisioned in the adoption of the "Americans with Disabilities Act" by the United States Congress several years ago.[106]

These observations are particularly apt in light of the Supreme Court of California's ruling that hereditary deafness was a defect of sufficient severity to recognize a child's claim for wrongful life.[107] Although the court did not award general damages on this claim,[108] the implications of its holding—and its implicit judgment that rather than suffer from deafness, the child should not have been born at all—are far-reaching in terms of our society's conceptions of disability.

Wrongful birth and wrongful life claims are particularly agonizing for the individual children at the center of the claims. Scholars have commented that

---

consistent with Alexander v. Choate, 469 U.S. 287 (1985)"); *cf.* 42 U.S.C. § 12112(b) (construing discrimination to include procedures and policies which "have the effect of discrimination on the basis of disability").

105. Edmonds v. W. Pennsylvania Hosp. Radiology Assocs., 607 A.2d 1083, 1088 (Pa. 1992) (Wieand, J., concurring).

106. Grubbs *ex rel.* Grubbs v. Barbourville Family Health Ctr., P.S.C., 120 S.W.3d 682, 692-93 (Ky. 2003) (Wintersheimer, J., concurring) (citation omitted).

107. *See* Turpin v. Sortini, 643 P.2d 954, 966 (Cal. 1982).

108. *Id.*

"both claims rely on a fundamental jurisprudential position that a life itself, if defective, can be a legal wrong, a 'damage' to the parents of that life and the child who lives it."[109] To establish causation for a wrongful birth or wrongful life claim, a mother must testify that she would not have chosen to carry the child to term if she had been informed of the defect in a timely manner. Such testimony is emotionally crippling not only to the child suffering from physical or mental infirmities, but also to the larger disability community that seeks equality of opportunity and full participation, both of which are goals supported by the ADA.

Wrongful birth and wrongful life claims are demoralizing to the disability community because they convey the message that an individual with a disability is inherently deficient. The discriminatory effects of these causes of action have been camouflaged by the rhetoric of reproductive choice. Wrongful birth and wrongful life claims hold doctors liable for failing to inform patients of the *risk* of potential genetic defects. In *Gleitman*, there was a twenty percent risk that the child would develop birth defects because of the mother's contraction of German measles early in her pregnancy.[110] At trial, the defendant physician testified that some doctors would recommend an abortion in situations where one out of five infants would be born with some defect; but the physician stated that "he did not think it proper to destroy four healthy babies because the fifth one would have some defect."[111] By affirming the parents' rights to discriminate in their reproductive choices on the basis of disability, wrongful life and wrongful birth claims send a demoralizing message to the disability community. As one academic noted, "any benefits that wrongful birth and wrongful life actions secure for the individual plaintiff come at the cost of demeaning and demoralizing anti-therapeutic messages delivered to the community of people with disabilities and to greater society."[112]

Pennsylvania State Senator M. Joseph Rocks discussed the obstacles posed by wrongful birth and wrongful life claims and recognized that these claims interfere with the goal of achieving a society that is free from discrimination based on disability. He stated that the intent of legislation barring wrongful birth and wrongful life claims is

> to stop a court-engendered policy which views the birth of a child, be that child handicapped or otherwise, as a damaging event for which someone should be punished in order to prevent this quality of life ethic from becoming so pervasive that a handicapped child is routinely considered better off dead and of less value than what we would call "a normal child" and to

---

109. James Bopp, Jr., Barry A. Bostrom & Donald A. McKinney, *The "Rights" and "Wrongs" of Wrongful Birth and Wrongful Life: A Jurisprudential Analysis of Birth Related Torts*, 27 DUQ. L. REV. 461, 463 (1989).

110. Gleitman v. Cosgrove, 227 A.2d 689, 690 (N.J. 1967), *abrogated by* Berman v. Allan, 404 A.2d 8 (N.J. 1979).

111. *Gleitman*, 227 A.2d at 690-91.

112. Hensel, *supra* note 32, at 164.

> prevent the practice of medicine . . . from becoming coerced into accepting eugenic abortion as a condition for avoiding [such] lawsuits.[113]

Wrongful birth and wrongful life claims violate Title II's mandate by having a disparate impact on children with disabilities in a manner that is not justified by the provisions or goals of the ADA.[114]

3. Wrongful Birth and Wrongful Life Claims Constitute Discrimination by a Public Entity

Legislation and judicial common law that recognize and enforce claims for wrongful birth and wrongful life constitute discriminatory state action[115] and, therefore, come within the purview of Title II of the ADA. Because Title II only prohibits discrimination by a public entity, in order for there to be a statutory violation, there must be governmental action.[116] For all practical purposes, this requirement is identical to the requirement of state action under the Fourteenth Amendment.[117] As such, no matter how discriminatory or unjust, private acts cannot serve as the basis of a Title II (or Fourteenth Amendment) violation.

Laws enacted by state legislatures are quintessential state action.[118] Consequently, laws that affirmatively create a cause of action and statutory remedy for wrongful birth or wrongful life claims are actions by a public entity under the purview of Title II. As noted above,[119] no cause of action existed for wrongful birth or wrongful life claims at common law. Therefore, when a state

---

113. PA. LEGIS. J. 1961 (Mar. 22, 1988).

114. *Cf.* Schierberl v. Bird, 848 F. Supp. 303, 313-14 (D. N.H. 1994) (finding that allegations that the New Hampshire Division of Children and Youth Services segregated disabled children in institutions, which isolated them from non-disabled children and denied them services and placement opportunities comparable to those available to non-disabled children, were sufficient claims to withstand a motion to dismiss); Baby Neal v. Casey, 821 F. Supp. 320, 329-30 (E.D. Pa. 1993) (holding that genuine issues of material fact precluded summary judgment on section 504 and ADA discrimination claims by two children who were allegedly classified as "unadoptable" by the Philadelphia Department of Human Services because of their disability and were thus excluded from certain programs necessary for adoption).

115. 15 AM. JUR. 2D *Civil Rights* § 10 (2000) ("State action includes action by a state legislature [and] state courts . . . .").

116. 42 U.S.C. § 12131(1)(A) (2000) (stating that "public entity" includes "any [s]tate or local government").

117. Section 1 of the Fourteenth Amendment requires that the wrongful conduct of private individuals have some connection to state authority to be actionable under the Fourteenth Amendment. *See* Civil Rights Cases, 109 U.S. 3, 17 (1883).

118. LAURENCE H. TRIBE, AMERICAN CONSTITUTIONAL LAW, § 18-1, at 1688-89 (2d ed. 1988); *see* Henry C. Strickland, *The State Action Doctrine and the Rehnquist Court*, 18 HASTINGS CONST. L.Q. 587, 599 (1991).

119. See *supra* Part I.B.

legislature enacts positive law, which creates a new right and a remedy for violation of that right, the state has unquestionably *acted.* Despite this recognition, a state court has held that a statute *prohibiting* wrongful birth or wrongful life claims does not amount to state action because the state was not responsible for the specific conduct of which the plaintiff complained; that is, the statute did "not encourage physicians to engage in the negligent conduct at issue."[120] This analysis is questionable in light of a subsequent federal district court opinion, which criticized this reasoning, acknowledging that the trial court's use of a state statute to bar the plaintiff's claim could have been "challengeable state action."[121] Consequently, state prohibitions of wrongful life and wrongful birth claims equally constitute state action.[122]

Likewise, judicial common law recognition of a tort cause of action constitutes government action. According to Laurence H. Tribe, "The general proposition that common law is state action—that is, that the state 'acts' when its courts create and enforce common law rules—is hardly controversial."[123] In particular, Tribe notes that tort law is comprised of a body of "government-made rules."[124] Similarly, in commenting on the state action requirement of the Fourteenth Amendment, the Supreme Court has held that "judicial action is not immunized from the operation of the Fourteenth Amendment simply because it is taken pursuant to the state's common-law policy."[125] Thus, if a state decides that the denial of an informed reproductive choice is a cognizable injury at law with compensable remedies, this decision constitutes state action.

Some may argue that in wrongful birth and wrongful life claims, the parents or the children themselves, rather than the state, are the ones who discriminate against the child with genetic abnormalities. While appealing on the surface, this argument lacks a strong foundation. Although the discrimination may appear to be the result of the parents' independent reproductive choices, the discrimination bears the heavy imprint of the state. Indeed, state law causes of action for wrongful birth and wrongful life would not exist without significant involvement from state governments in

120. Edmonds v. W. Pennsylvania Hosp. Radiology Assocs., 607 A.2d 1083, 1087-88 (Pa. 1992).

121. Flickinger v. Wanczyk, 843 F. Supp. 32, 37 n.3 (E.D. Pa. 1994).

122. But even if statutes prohibiting wrongful birth and wrongful life claims do not amount to state action, one could not conclude that state laws recognizing wrongful birth and wrongful life claims also do not amount to state action. At common law, no cause of action existed for wrongful birth or wrongful life; hence, the recognition of such claims by the state through legislation or through judicial common law necessarily represents an affirmative change in the status quo, which would be deemed state *action.* In contrast, the prohibition of such claims is best characterized as preserving the status quo, and consequently state *inaction.* The Supreme Court has recognized that "the mere denial of judicial relief" for certain private acts does not convert those private acts into state action. Flagg Bros., Inc. v. Brooks, 436 U.S. 149, 165 (1978).

123. TRIBE, *supra* note 118, § 18-6, at 1711.

124. *Id.* at 1713.

125. Shelley v. Kraemer, 334 U.S. 1, 20 (1948).

recognizing and enforcing such claims.[126] Moreover, the parents' decision to discriminate against their children with potential disabilities by refusing to carry the offspring to term is not actionable under the ADA. Title II's "public entity" requirement means that parents may make discriminatory reproductive decisions based on eugenic considerations without running afoul of the ADA. Rather, the conduct that is offensive to the ADA is the *state's* decision to make the parents' discriminatory reproductive decisions a legal right to which they are entitled compensation if deprived. Accordingly, wrongful birth and wrongful life claims constitute "a community pronouncement, via a *government* institution, that an individual's life with impairments is worse than nonexistence, or that a reasonable person would have aborted a now-living child."[127]

Courts have found the existence of state action by private individuals or entities who commit the challenged activity if there is a "sufficiently close nexus" between the action and the state.[128] In an analogous situation, when determining whether a patient's right to refuse life-sustaining treatment can be premised on constitutional rights, courts have found the state action sufficient to invoke the constitutional right to privacy.[129] For example, the Supreme Court of Washington has held that

> the presence of the state is manifested by its capability of imposing criminal sanctions on the hospital and its staff, by its licensing of physicians, by the required involvement of the judiciary in the guardianship appointment process, and by the State's parens patriae responsibility to supervise the affairs of incompetents.[130]

Similarly, the state judiciary necessarily must be involved in recognizing, adjudicating, and enforcing wrongful birth and wrongful life causes of action, as well as protecting incompetents. Given the state's required involvement in such affairs, wrongful birth and wrongful life claims involve sufficient state action. Thus, there can be no question that the state has acted when the challenged action is taken by the state legislature or judiciary via recognition of wrongful birth and wrongful life causes of action.

---

126. *See id.* at 19 (holding that state court enforcement of racially restrictive covenants constitutes state action under the Fourteenth Amendment because "[i]t is clear that but for the active intervention of the state courts, . . . [purchasers] would have been free to occupy the properties in question without restraint").

127. Hensel, *supra* note 32, at 173 (emphasis added).

128. Jackson v. Metro. Edison Co., 419 U.S. 345, 351 (1974), *quoted in* Eichner v. Dillon, 426 N.Y.S.2d 517, 540 (N.Y. App. Div. 1980); *In re* Colyer, 660 P.2d 738, 742 (Wash. 1983).

129. *See Eichner*, 426 N.Y.S.2d at 540; *In re* Colyer, 660 P.2d at 742. *But see* Schiavo *ex rel.* Schindler v. Schiavo, 357 F. Supp. 2d 1378, 1388 (M.D. Fla. 2005).

130. *In re* Colyer, 660 P.2d at 742 (citations omitted). The court added that "[t]his privacy right, if founded on the federal constitution and applied to the states through the Fourteenth Amendment, extends only to situations where state action exists." *Id.*

Some might argue that finding state action in this context would alter the balance of federalism or lead to the federalization of all state-created rights, but this "slippery-slope" argument ignores the limiting principle of discriminatory state action. Although states remain free to create the rights and remedies they deem appropriate for their citizens, they cannot unreasonably discriminate against a class of individuals. As discussed earlier in this Article,[131] only state action that discriminates against individuals on the basis of disability is subject to the federal mandate of Title II. Thus, prohibiting state action that discriminates against individuals with disabilities is no more of an encroachment on state autonomy than are federal efforts to remedy state discrimination on other categories, such as race. Accordingly, if the ADA's twin goals of equal opportunity and full participation of individuals with disabilities are to be realized, wrongful birth and wrongful life claims have no place in our society.[132]

III. BEYOND THE WRONGFUL RHETORIC

Because wrongful birth and wrongful life claims contravene the provisions of Title II of the ADA, it is necessary to examine the purposes these two causes of action serve and explore how these purposes can be addressed through non-discriminatory alternatives that are consistent with the goals of the ADA. As tort claims, generally wrongful birth and wrongful life causes of action serve two primary purposes: (1) they establish a means to deter negligent conduct by health care providers; and (2) they compensate parents and children for injuries or losses sustained as result of negligent conduct.

A. Deterring Negligent Conduct Through Other Avenues

One of the primary functions of wrongful birth and wrongful life claims is to deter negligent conduct in pre-conception genetic testing and counseling and prenatal testing. As a preliminary note, medical and scientific advancements in the fields of genetics and fetal surgery have the potential to transform wrongful birth and wrongful life claims into traditional negligence suits.[133] With the

---

131. *See supra* Part II.B.3.

132. A state law cause of action for wrongful birth or wrongful life may be preempted by the ADA because it, much like other state laws, "stands as an obstacle to the accomplishment and execution of the full purposes and objectives of Congress." Hines v. Davidowitz, 312 U.S. 52, 67-68 (1941) (holding that federal alien registration law precluded enforcement of state law on the same subject, especially where legislation dealt with "the rights, liberties, and personal freedom of human beings, and is in an entirely different category from state tax statutes or state pure food laws"). Wrongful birth and wrongful life claims stand as an obstacle to Congress' purposes in enacting and enforcing federal antidiscrimination disability laws. Consequently, wrongful birth and wrongful life claims must be invalidated to the extent that they conflict with the objectives of federal anti-discrimination laws, such as the ADA.

133. Thomas A. Warnock, *Scientific Advancements: Will Technology Make the Unpopular*

ever-increasing ability to correct birth defects in utero, a physician's failure to diagnose or detect a congenital birth defect will be deemed to cause the birth defect under traditional principles of proximate causation.

Nevertheless, tort claims are not the only means of addressing health care malpractice. In fact, there is little evidence to show that malpractice suits reduce the incidence of negligence; scholars often bemoan the "poor fit" between the incidence of medical negligence and the incidence of lawsuits.[134] Moreover, "only a few patients injured by medical malpractice ever actually sue, and because 'many malpractice lawsuits are brought and won by patients even though expert reviewers can identify no evidence of negligent care,' there is little evidence that such suits provide any real deterrence to future acts of negligence."[135] Furthermore, allowing tort claims like wrongful birth and wrongful life may encourage physicians to practice defensive medicine, which could lead to increases in the cost of healthcare. For example, a physician practicing defensive medicine will order all available diagnostic tests in an effort to avoid liability for potentially exorbitant compensatory damages, regardless of the cost or need for the diagnostic test.

Aside from tort claims, several other mechanisms can redress medical negligence. One option is to increase specialized training and continuing education in the field of genetic testing. Such training should become mandatory for obstetricians, gynecologists, and primary care physicians involved in prenatal testing. Inadequate education and training "in the fast-changing field of genetics" is a major problem in the area of prenatal care, which is increasingly tied to genetics.[136] Another effective alternative to reducing medical negligence is disciplinary action by state medical licensing boards. Furthermore, increased state and federal regulation of laboratories that conduct genetic testing may lead to reductions in medical negligence. Therefore, though they are imperfect mechanisms, increased education and training, administrative remedies such as disciplinary proceedings, and increased regulation of genetic testing may aid in deterring medical negligence without discriminating against individuals with disabilities and thereby violating Title II's mandate.

Wrongful Birth/Life Causes of Action Extinct?, 19 TEMP. ENVTL. L. & TECH. J. 173, 184 (2001) (arguing that extremely rapid scientific advancements in the field of genetic engineering and fetal surgery will allow physicians to detect and treat genetic defects in unborn fetuses and consequently permit plaintiffs to recover damages under existing causes of action).

134. Hensel, *supra* note 32, at 192.

135. *Id.* (quoting Michelle M. Mello & Troyen A. Brennan, *Deterrence of Medical Errors: Theory and Evidence for Malpractice Reform*, 80 TEX. L. REV. 1595, 1618-20 (2002)).

136. Mark A. Rothstein & Sharona Hoffman, *Genetic Testing, Genetic Medicine, and Managed Care*, 34 WAKE FOREST L. REV. 849, 858 (1999).

B. *The Wrongful Reasoning of Compensation*

The second goal of wrongful birth and wrongful life claims is to provide parents or guardians with an independent source of funding to cover extraordinary medical, educational, or other costs related to disability. Consequently, wrongful birth and wrongful life claims operate as a type of supplemental insurance that spreads such financial hardships to parties who are more likely to be able to bear the cost. But this goal is overstated by supporters of wrongful birth and wrongful life actions and it obscures the real effect of these claims in several ways.

First, as negligence claims, wrongful birth and wrongful life claims do not center their focus on the individual's need for monetary assistance. In terms of providing needed financial assistance to families with disabled children, wrongful birth and wrongful life claims are both overinclusive and underinclusive; they may provide substantial, but unnecessary, financial support to some families and provide no relief to other families with compelling need for financial support.

Second, wrongful birth and wrongful life claims do not compensate every individual who is deprived of the ability to make an "informed" reproductive choice. Due to the relaxed causation requirement discussed earlier in this Article,[137] recovery is limited to cases where parents testify that they would have aborted the child or where children testify that they should have been aborted. Financial "assistance is provided only to those willing to openly disavow their self-worth and dignity."[138] Nonetheless, not all harms are compensable, and the "miracle of life" is never a guarantee. All parents face the risk that their child will be born with genetic deficiencies. Parents who openly discount the worth of their children with disabilities should not be rewarded monetarily through tort claims that are unavailable to parents who accept and care for such children. By compensating those who devalue children with congenital defects and denying recovery to those who embrace these children, wrongful birth and wrongful life claims create a perverse system that sends a negative message about the value of life with disability.

C. *Reaffirming Life with Disability*

At least one author has pointed to the inconsistencies in judicial treatment of cases involving "quality of life" determinations, arguing that recognition of the right to die necessitates recognition of wrongful birth and wrongful life claims:

> It seems incongruous, then, that it is possible to make a choice between existence and non-existence by refusing to undergo [medical] treatment, yet this choice is refused to those who assert that they have been deprived of the

137. *See supra* Part I.A.

138. Hensel, *supra* note 32, at 171.

> opportunity of making this choice (albeit by proxy) by preferring not to have been born. If it is permissible to prefer death (non-existence) and exercise this preference by refusing treatment, with legal remedies available to back up attempts to thwart the exercise of this choice, why should it not also be permissible to prefer non-existence (abortion) and, since exercising this preference by choosing to be aborted rather than born has been denied the child, why should it not seek compensation?[139]

The greater contradiction, however, is in society's treatment of individuals at the beginning and end stages of life as contrasted to its treatment of individuals during the remainder of their lifetimes. Against the landscape of federal and state policies protecting citizens with disabilities, wrongful birth and wrongful life claims "create[] a jurisprudential contradiction: a legal system that at once encourages the destruction of 'defective' lives before birth and seeks to protect those lives after birth."[140] If consistency is to be the goal of law, then let it not be for its own sake, but let it foster respect for the inherent worth of every individual, including individuals with disabilities, at each and every stage of their lives.

---

139. Beaumont, *supra* note 9, at 109.

140. Bopp, Bostrom & McKinney, *supra* note 109, at 495 (citing Victor G. Rosenblum & Edward R. Grant, *The Legal Response to Babies Doe: An Analytical Prognosis*, 1 ISSUES IN L. & MED. 391 (1986)).

WHOSE LAW IS IT ANYWAY? THE CULTURAL LEGITIMACY OF INTERNATIONAL HUMAN RIGHTS IN THE UNITED STATES

ELIZABETH M. BRUCH*

"The Court's discussion of these foreign views . . . is therefore meaningless dicta. Dangerous dicta, however, since 'this Court . . . should not impose foreign moods, fads, or fashions on Americans.'" - Justice Antonin Scalia (2003).[1]

"It does not lessen our fidelity to the Constitution or our pride in its origins to acknowledge that the express affirmation of certain fundamental rights by other nations and peoples simply underscores the centrality of those same rights within our own heritage of freedom." - Justice Anthony Kennedy (2005).[2]

"We have our own law, we have our own traditions, we have our own precedents, and we should look to that in interpreting our Constitution." - Justice Samuel Alito (2006).[3]

INTRODUCTION

The relationship between the United States and international law is complex and evolving, implicating political, cultural, and legal considerations. In the realm of international human rights law, the relationship is particularly and deeply contested. In recent years, the bounds of that relationship have been challenged in the national media, in public discourse, and in all three branches

---

* Associate Professor of Law, Valparaiso University School of Law. I wish to thank Abdullahi A. An-Na'im, Charles Howard Candler Professor of Law at Emory University School of Law, for his thoughtful comments on an early draft of this article. In addition, I am grateful to Johanna Bond, Matt Fraidan, Vivian Hamilton, Margaret Johnson, Kate Kruse, and Vicki Philips (the Potomac Valley Writers Workshop) for their valuable support and critique at various stages in my writing process, to my colleagues at Valparaiso, and to Chris Johnson for early discussions and comments on this article. I also appreciate the research and editing assistance provided by Leslie Wilson, Aimee Brown, Theresa Ellis, and Regina Reynolds. Finally, I am thankful to Dave Gage for his perspective and encouragement.

1. Lawrence v. Texas, 539 U.S. 558, 598 (2003) (Scalia, J., dissenting) (quoting Foster v. Florida, 537 U.S. 990, 990 n.* (2002) (Thomas, J., concurring in denial of certiorari)).

2. Roper v. Simmons, 543 U.S. 551, 578 (2005).

3. *Confirmation Hearing on the Nomination of Samuel A. Alito, Jr. to be an Associate Justice of the Supreme Court of the United States: Hearing Before the Comm. on the Judiciary*, 109th Cong. 370 (2006) (statement of Judge Samuel A. Alito), *available at* http://www.washingtonpost.com/wp-dyn/content/article/2006/01/10/AR2006011000781.html [hereinafter *Alito Confirmation Hearing*].

of the federal government.[4] In an era where the public focuses much of its attention on the appropriate role of the judiciary, a series of United States Supreme Court decisions that cite to international human rights law have been controversial. In its 2002 decision in *Atkins v. Virginia*,[5] its 2003 decision in *Lawrence v. Texas*,[6] and its 2005 decision in *Roper v. Simmons*,[7] the Supreme Court invoked international human rights law and the practice of other nations in resolving issues of fundamental rights to life and privacy.

The Court's action has drawn both praise and vilification and has invited serious inquiry into the legitimacy of such a judicial approach. Most scholars who have undertaken this inquiry have considered the Supreme Court's conduct from either a perspective of comparative constitutionalism,[8] or as a

---

4. *See* Charles Lane, *The Court is Open for Discussion; AU Students Get Rare Look at Justices' Legal Sparring*, WASH. POST, Jan. 14, 2005, at A1 (recounting the debate between Justice Antonin Scalia and Justice Stephen G. Breyer at American University Washington College of Law over the Supreme Court's use of foreign and international law in its decisions); Tony Mauro, *U.S. Supreme Court vs. The World*, USA TODAY, June 20, 2005, at 15A, *available at* 2005 WLNR 9728892 (reviewing the current controversy over the use of international law by the Supreme Court); Abdon M. Pallasch, *Justices Shouldn't Cite Foreign Laws, U.S. Attorney General Tells Students*, CHICAGO SUN-TIMES, Nov. 10, 2005, at 24, *available at* 2005 WLNR 19361771 (reporting Attorney General Alberto Gonzales's remarks against the citation of "foreign law" in Supreme Court opinions).

5. 536 U.S. 304 (2002) (prohibiting the execution of the developmentally disabled). Although the Court used the descriptive term of "mentally retarded," which was the language used in the earlier proceedings, this Article will instead use the term "developmentally disabled" to refer to the individuals whose rights are at stake in the decision.

6. 539 U.S. 558 (2003) (finding a criminal sodomy statute unconstitutional).

7. 543 U.S. 551 (2005) (prohibiting the execution of juvenile offenders).

8. A comparative constitutionalist perspective considers the Supreme Court's citations to foreign and international sources principally in determining when, if at all, it is appropriate for a domestic court interpreting its national constitution to look to sources outside that constitution framework. *See, e.g.*, Roger P. Alford, *Federal Courts, International Tibunals, and the Continuum of Deference: A Postscript on Lawrence v. Texas*, 44 VA. J. INT'L L. 913, 926-28 (2004); Kenneth Anderson, *Foreign Law and the U.S. Constitution*, 131 POL'Y REV. 33, 34 (June – July 2005); William N. Eskridge, Jr., *United States:* Lawrence v. Texas *and the Imperative of Comparative Constitutionalism*, 2 INT'L J. CONST. L. 555, 556 (2004); Eric A. Posner, *Transnational Legal Process and the Supreme Court's 2003-2004 Term: Some Skeptical Observations*, 12 TULSA J. COMP. & INT'L L. 23, 25, 36-37 (2004); John Yoo, *Peeking Abroad?: The Supreme Court's Use of Foreign Precedents in Constitutional Cases*, 26 U. HAW. L. REV. 385, 393 (2004). Scholars distinguish between "soft" and "hard" use of foreign and international legal sources, including empirical use (evidence), reason-borrowing, normative use (moral-factfinding), and use as a matter of comity. Taavi Annus, *Comparative Constitutional Reasoning: The Law and Strategy of Selecting the Right Arguments*, 14 DUKE J. COMP. & INT'L L. 301, 312 (2004) (describing as "soft" the use of foreign materials without relying on them in reaching the holding and "hard" use as the direct contribution of comparative materials in the holding either through normative reasoning or empirical reasoning). Some scholars are critical of all uses, others of "hard" uses. *Id.*; *see also* Alford, *supra*, at 926-27 (using the "continuum of deference" model, which is used to distinguish between the varying

matter of the domestic implementation of international law.[9] As an alternative to those perspectives, this Article examines the national controversy over the Supreme Court's decisions in *Atkins*, *Lawrence*, and *Simmons* through the lens of international human rights law. When states undertake international human rights obligations, they become responsible for ensuring compliance with those obligations within their domestic legal orders.[10] This state responsibility includes punishing violations of human rights, as well as respecting and ensuring the free exercise of those rights.[11] Governments are expected to take

---

degrees of deference to foreign materials, and generally opposing U.S. deference to these materials); Eskridge, *supra*, at 556-59 (approving use of foreign materials as indicative of an "emerging normative consensus," as a matter of "international comity," and as a recognition of "pluralism"); Joan L. Larsen, *Importing Constitutional Norms from a "Wider Civilization": Lawrence and the Rehnquist Court's Use of Foreign and International Law in Domestic Constitutional Interpretation*, 65 OHIO ST. L.J. 1283, 1298 (2004) (criticizing domestic use of foreign materials unless there is adequate justification for its use); Yoo, *supra*, at 393 (opposing U.S. deference to foreign materials generally).

9. The main issue from this perspective is whether international law has become binding at the domestic level; of course, this would never be an issue regarding foreign law sources. *See, e.g.*, Diane Marie Amann, *"Raise the Flag and Let It Talk": On the Use of External Norms in Constitutional Decision Making*, 2 INTL. J. CONST. L. 597, 608-09 (2004).

10. The major human rights treaties contain general language of obligation to this effect. For example, Article 2 of the International Covenant on Civil and Political Rights provides:

> 1. Each State Party to the present Covenant undertakes to respect and to ensure to all individuals within its territory and subject to its jurisdiction the rights recognized in the present Covenant, without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status.
> 2. Where not already provided for by existing legislative or other measures, each State Party to the present Covenant undertakes to take the necessary steps, in accordance with its constitutional processes and with the provisions of the present Covenant, to adopt such legislative or other measures as may be necessary to give effect to the rights recognized in the present Covenant.
> 3. Each State Party to the present Covenant undertakes:
> (a) To ensure that any person whose rights or freedoms as herein recognized are violated shall have an effective remedy, notwithstanding that the violation has been committed by persons acting in an official capacity;
> (b) To ensure that any person claiming such a remedy shall have his right thereto determined by competent judicial, administrative or legislative authorities, or by any other competent authority provided for by the legal system of the State, and to develop the possibilities of judicial remedy;
> (c) To ensure that the competent authorities shall enforce such remedies when granted.

International Covenant on Civil and Political Rights, G.A. Res. 2200A, at 2, U.N. GAOR, 21st Sess., Supp. No. 16, U.N. Doc. A/6316 (Dec. 16, 1966), *available at* http://www.ohchr.org/english/law/ccpr.htm [hereinafter ICCPR].

11. *Id.*; *see also Velásquez Rodríguez Case*, 1988 Inter-Am. C.H.R. (ser. C) No. 4, at ¶¶ 164-65 (July 29, 1988) (The obligation of states to ensure free and full exercise of human rights "implies the duty of States Parties to organize the governmental apparatus and, in general, all the structures through which public power is exercised, so that they are capable of juridically ensuring the free and full enjoyment of human rights.").

the lead in creating domestic conditions that foster fulfillment of their international human rights obligations. The relative ease or difficulty of this implementation process depends largely on the degree to which the international standards comport with existing domestic values.[12]

Although the United States has often taken the lead globally in advancing international human rights, it has been notoriously reluctant to ratify international human rights agreements or to incorporate those agreements into domestic law.[13] Some of this reluctance has stemmed from a U.S. perception that domestic law already provides significant protection for individual rights.[14] This sense of international leadership and dominance persists in the current resistance to international law, but it has been increasingly overshadowed by concerns grounded in the converse notion of isolation and fear that other nations are now mandating standards for the United States.[15] Despite frequent

---

12. Harold Hongju Koh, *Why Do Nations Obey International Law?*, 106 YALE L.J. 2599, 2645-46 (1997) (discussing the theory that internalized compliance and obedience will increase comportment with international law); *see* Abdullahi A. An-Na'im, *State Responsibility Under International Human Rights Law to Change Religious and Customary Laws*, *in* HUMAN RIGHTS OF WOMEN 167, 169 (Rebecca J. Cook ed., 1994) [hereinafter An-Na'im, *State Responsibility*] ("In practice, a state's willingness or ability to influence practices based on religious and customary laws depends on many factors, any of which could cause difficulty in situations where domestic religious and customary laws are likely to be in conflict with internationally recognized standards of human rights.").

13. The current Bush administration is well-known not only for its reluctance to undertake international obligations but also for its willingness to disavow them. *See* Neil A. Lewis, *U.S. Rejects All Support for New Court on Atrocities*, N.Y. TIMES, May 7, 2002, at A11, *available at* 2002 WLNR 4088405 (documenting the Bush administration's "unsigning" of the International Criminal Court treaty, originally signed by President Clinton); *U.S. Won't Follow Climate Treaty Provisions, Whitman Says*, N.Y. TIMES, Mar. 28, 2001, at A19 (reporting that the Bush administration would not support the Kyoto international climate treaty). While previous administrations have been more willing to undertake international obligations, they frequently placed conditions on those obligations or faced resistance from Congress. *See* Connie de la Vega, *Human Rights and Trade: Inconsistent Application of Treaty Law in the United States*, 9 UCLA J. INT'L L. & FOREIGN AFF. 1, 11-13 (2004) (criticizing the United States' history of conditional ratification); David Sloss, *The Domestication of International Human Rights: Non-Self-Executing Declarations and Human Rights Treaties*, 24 YALE J. INT'L L. 129, 139-42 (1999) (describing the history of ratification of human rights treaties in the United States).

14. *See, e.g.*, Abdullahi A. An-Na'im, *Round Table Discussion on International Human Rights Standards in the United States: The Case of Religion or Belief*, 12 EMORY INT'L L. REV. 973, 977 (1998) [hereinafter An-Na'im, *Round Table Discussion*]. The remarks of participant Jeremy Gunn illustrate this idea: "[A]lthough the United States does not apply the international standards to itself, it nevertheless—with important exceptions—generally acts in accordance with international standards and often exceeds them." *Id.* Another participant, David Bederman, echoed this perception of a "culture of compliance" but noted the risks of the United States' sense of constitutional exceptionalism. *Id.* at 982.

15. *See supra* note 4; *see also* Phyllis Schlafly, *Is Relying on Foreign Law an Impeachable Offense?*, EAGLE FORUM, Mar. 16, 2005, http://www.eagleforum.org/column/2005/mar05/05-03-16.html (expressing outrage that the

claims of "American exceptionalism," the United States is currently engaging in the same struggle that other nations undergo in the process of reconciling international legal standards and obligations with domestic cultural values and policy concerns. The United States Supreme Court has been, and continues to be, an important part of that process by playing an appropriate, though limited, role in advancing the cultural legitimacy of international human rights in the United States.

The national controversy about the Supreme Court's decisions has been misguided in several respects that merit brief clarification. As an initial point, the public discourse on the issue regularly, and incorrectly, conflates international law[16] with "foreign law," which is the domestic law of other nations.[17] Moreover, the use of these collective "foreign" sources is often characterized in a way that suggests that the Court has treated the sources as binding or otherwise authoritative.[18] In reality, the Supreme Court has cited foreign law[19] and international law[20] in its decisions, but the Court has never

---

Supreme Court used foreign law and international opinion in its *Simmons* decision).

16. Three sources comprise international law: treaties, customary law, and general principles of law. RESTATEMENT (THIRD) OF FOREIGN RELATIONS LAW OF THE UNITED STATES § 102 (1987). For example, in *Roper v. Simmons*, when the Supreme Court cited to the United Nations Convention on the Rights of the Child, the Court called it an international authority. 543 U.S. 551, 575-76 (2005). When the United States is party to a treaty, that treaty is binding on the United States as an international legal obligation. RESTATEMENT, *supra*, § 111. The treaty is enforceable in U.S. courts if it is self-executing or if it has been implemented by domestic legislation. *Id.* Customary international law is generally binding on all states. *Id.* §§ 701-02.

17. Foreign law sources would include the laws of other countries or the decisions of their national courts. For example, in *Lawrence v. Texas*, the Court cites foreign law when it cites to the criminal legislation of the United Kingdom. 539 U.S. 558, 572-73 (2003). In conflating the two, many have referred to both international law and foreign law as "foreign" sources, with the pejorative connotations of that term. *See, e.g.*, *infra* Part II.C. Foreign law, of course, would not be binding on a U.S. court (although parties to litigation in a U.S. court may be bound in some circumstances by foreign law).

18. *Confirmation Hearing on the Nomination of John G. Roberts, Jr. to be Chief Justice of the United States: Hearing Before the Comm. on the Judiciary*, 109th Cong. 293 (2005), *available at* http://www.washingtonpost.com/wp-dyn/content/article/2005/09/14/AR2005091401445.html [hereinafter *Roberts Confirmation Hearing*] (Question by Senator Coburn: "My question to you is, relying on foreign *precedent* and selecting and choosing a foreign *precedent* to create a bias outside of the laws of this country, is that good behavior?" (emphasis added)); *see also Alito Confirmation Hearing*, *supra* note 3, at 370 (Question by Senator Kyl: "What is the proper role, in your view, of foreign law in U.S. Supreme Court decisions, and when, if ever, is citation to or *reliance* on these foreign laws appropriate?" (emphasis added)).

19. For example, in *Lawrence*, the Court cited to British law repealing the criminal laws punishing homosexual conduct. *Lawrence*, 539 U.S. at 572-73 (citing the United Kingdom's Sexual Offences Act 1967, c. 60 § 1). Similarly, in *Simmons*, the Court cited British laws prohibiting the execution of juveniles. *Simmons*, 543 U.S. at 577 (citing the Children and Young Person's Act, 1933, 23 Geo. 5, c. 12 (U.K.), and the Criminal Justice Act, 1948, 11 & 12

suggested that the foreign or international law sources it cited were binding.[21] The *Atkins*, *Lawrence*, and *Simmons* decisions are firmly grounded in domestic U.S. law.[22]

So why is there a controversy at all? In the current international legal system, the formation and development of international law, including international human rights law, occurs primarily at the international level through the creation of bilateral and multilateral treaties and through interpretive mechanisms such as treaty-bodies and international tribunals. In contrast, the implementation of international human rights law is primarily entrusted to the national or domestic front.[23] Through their governments, nation-states are the primary actors in developing and implementing international human rights law at both the international and national levels.[24]

Geo. 6, c. 58 (U.K.)).

20. The Court cited numerous international human rights treaties in *Simmons*—including the United Nations Convention on the Rights of the Child, the International Covenant on Civil and Political Rights, the American Convention on Human Rights, and the African Charter on the Rights and Welfare of the Child. *Simmons*, 543 U.S. at 576. In *Lawrence*, the Court cited a decision of the European Court of Human Rights. *Lawrence*, 539 U.S. at 573 (citing Dudgeon v. United Kingdom, 45 Eur. Ct. H.R. ¶ 52 (1981)).

21. In its decisions, the Court has often taken pains to emphasize this. *See infra* Part II.

22. *See* discussion of decisions *infra* Parts II.B-C.

23. The international mechanisms to monitor states' compliance with international human rights obligations include both "treaty bodies," such as the Human Rights Committee (created to monitor compliance with the International Covenant on Civil and Political Rights), and non-treaty bodies, such as the Human Rights Commission, now the Human Rights Council (a subsidiary body of the United Nations to monitor general compliance with international human rights obligations). But these international mechanisms are generally considered to be avenues of last resort. *See, e.g.*, ICCPR, *supra* note 10, at art. 41(c) (requiring exhaustion of all available domestic remedies before the Committee will consider a matter); *see also* Joan Fitzpatrick, *The Role of Domestic Courts in Enforcing International Human Rights Law*, *in* GUIDE TO INTERNATIONAL HUMAN RIGHTS PRACTICE 247, 261-62 (Hurst Hannum ed., 3d ed. 1999) (describing the primacy of national mechanisms in protecting human rights). There is a presumption in favor of resolving problems at the national level, relying on national governments to set up their domestic systems to protect human rights, provide remedies for violations, and generally implement international human rights obligations. The primary role of the international mechanisms is to establish and develop normative standards.

24. International human rights law is essentially a modern, post-World War II phenomenon. Richard B. Bilder, *An Overview of International Human Rights Law*, *in* GUIDE TO INTERNATIONAL HUMAN RIGHTS PRACTICE, *supra* note 23, at 3, 4-5. The United States and the other victorious Allied Nations played a foundational role in the development of both the substantive components of international human rights law and the international monitoring mechanisms. *Id.* United States historians note the important role played by Eleanor Roosevelt and others in developing the Universal Declaration of Human Rights, the foundational human rights document, which was later "codified" into the International Covenant on Civil and Political Rights and the International Covenant on Economic, Social and Cultural Rights. *See* John P. Humphrey, HUMAN RIGHTS AND THE UNITED NATIONS: A GREAT ADVENTURE 31-33, 42-43 (1984); John P. Humphrey, *The Universal Declaration of Human Rights: Its History, Impact*

Nonetheless, there is often popular concern that international human rights law at the domestic level represents the view of outsiders "imposed" on the nation, and that its domestic use offends national sovereignty and is sure to be out of step with domestic views and values. In the United States, this concern manifests itself in a wide range of contexts: from the popularity of "Get the US out of the UN" bumper stickers,[25] or the dismissive presidential rhetoric and decision-making,[26] to the proposed Congressional "Reaffirmation of American Independence Resolution."[27] Most immediately, concerns about international law are threaded throughout the current controversy over the role of the judiciary and the concern that "activist judges" will influence or change domestic law in ways that are contrary to domestic values.[28]

Concerns about judicial activism and the imposition of "foreign" values are not unique to the United States or to the area of international human rights law;[29] they commonly appear in the human rights arena as an aspect of the

---

and Juridical Character, in HUMAN RIGHTS: THIRTY YEARS AFTER THE UNIVERSAL DECLARATION 21-37 (B.G. Ramcharan ed., 1979). Nonetheless, many legal, philosophical, religious, and moral systems find support for the underlying principles of fundamental human rights in their own traditions and historical roots. *See, e.g.*, African Charter on Human and Peoples' Rights pmbl., June 27, 1981, 21 I.L.M. 58 [hereinafter African Charter].

25. *See, e.g.*, Cafepress.com, US out of UN – UN out of US!, http://www.cafepress.com/unwelcome (last visited Jan. 12, 2007); *see also* GetUSout.org, Get US out! of the United Nations, http://www.getusout.org/ (last visited Jan. 12, 2007) (describing the campaign of the John Birch Society to get the United States out of the United Nations).

26. The current administration has frequently expressed dissatisfaction and frustration with the United Nations. Many view the recent appointment of John Bolton as the U.S. Representative to the United Nations as a further indication of this dissatisfaction. Bolton is notorious for his anti-U.N. comments; Bolton has stated that "if 10 floors of the 38-story U.N. headquarters building were eliminated, 'it wouldn't make a bit of difference.'" Charles Babington & Dafna Linzer, *Bolton Assures Senators of Commitment to U.N.*, WASH. POST, Apr. 12, 2005, at A1, A10. Bolton has also stated, "There is no United Nations. There is an international community that occasionally can be led by the only real power left in the world—that's the United States—when it suits our interests and when we can get others to go along." Editorial, *Questioning Mr. Bolton*, N.Y. TIMES, Apr. 13, 2005, at A18; *see also* Christian Bourge, *Analysis: Bolton Controls Rhetoric*, UNITED PRESS INT'L, Apr. 11, 2005 ("Bolton has a long history of criticizing multinational institutions, and the United Nations in particular . . . .").

27. The Resolution, proposed by U.S. Rep. Feeney and others, states:

> That it is the sense of the House of Representatives that judicial interpretations regarding the meaning of the Constitution of the United States should not be based in whole or in part on judgments, laws, or pronouncements of foreign institutions unless such foreign judgments, laws, or pronouncements inform an understanding of the original meaning of the Constitution of the United States.

H.R. Res. 97, 109th Cong. (2005).

28. *Alito Confirmation Hearing*, *supra* note 3, at 410 (statement of Senator Sessions: "We believe that there has been a liberal social agenda being promoted too often by the courts that is foreign to our history and contrary to the wishes of the American people."); *see supra* notes 4, 15.

29. For example, in the negotiations of the North American Free Trade Agreement, state

debate between universalism and cultural relativism. Most international human rights advocates, and many of the human rights instruments themselves, assert the universality of human rights standards.[30] Skeptics often reject these sweeping claims in the name of "culture" and cultural relativism; similar arguments have been raised by governments that seek to avoid international human rights obligations,[31] and increasingly by a wide spectrum of activists, policymakers, and scholars who challenge the hegemony of the "western" view of international human rights.[32] In its relatively brief history, international

---

and national authorities were concerned about submitting to a "supranational" authority. *See* Stephen Zamora, *Allocating Legislative Competence in the Americas: The Early Experience Under NAFTA and the Challenge of Hemispheric Integration*, 19 HOUS. J. INT'L L. 615, 633 (1997).

30. "Whereas recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world . . . ." Universal Declaration of Human Rights, G.A. Res. 217A, at 71, U.N. GAOR, 3d Sess., Supp. No. 16, U.N. Doc. A/810 (Dec. 12, 1948) [hereinafter UDHR]; "Considering the obligation of States under the Charter of the United Nations to promote universal respect for, and observance of, human rights and freedoms . . . ." International Covenant on Economic, Social and Cultural Rights, G.A. Res. 2200A, at 49, U.N. GAOR, 21st Sess., Supp. No. 16, at 49, U.N. Doc. A/6316 (Dec. 16, 1966) [hereinafter ICESCR]. This claim of universality is derived from the very nature of human rights, which are premised on a common, shared humanity and recognition of entitlement to basic human dignity that is equally available to all, regardless of distinctions such as race, gender, religion, and other fundamental characteristics. In 1993, the principle of universality was explicitly re-embraced at the United Nations World Conference on Human Rights:

> The World Conference on Human Rights reaffirms the solemn commitment of all States to fulfil [sic] their obligations to promote universal respect for, and observance and protection of, all human rights and fundamental freedoms for all in accordance with the Charter of the United Nations, other instruments relating to human rights, and international law. The universal nature of these rights and freedoms is beyond question.

World Conference on Human Rights, June 14-25, 1993, *Vienna Declaration and Programme of Action*, ¶ I.1, U.N. Doc. A/CONF.157/23 (July 12, 1993).

31. *See* Michael C. Davis, *Human Rights in Asia: China and the* Bangkok Declaration, 2 BUFF. J. INT'L L. 215, 226-27 (1996) (quoting a speech by Liu Huaqiu, Head of the Chinese Delegation to the World Conference on Human Rights in Vienna, criticizing infringements on sovereignty in the name of universal human rights).

32. The international community and its mechanisms are generally criticized for being dominated by western, developed nations, and further criticized because the membership of international bodies is overwhelmingly male, even when there is a broader geographical and racial distribution. *See, e.g.*, Abdullahi Ahmed An-Na'im, *Problems of Universal Cultural Legitimacy for Human Rights*, *in* HUMAN RIGHTS IN AFRICA: CROSS-CULTURAL PERSPECTIVES 331, 348-53 (Abdullahi Ahmed An-Na'im & Francis M. Deng eds., 1990) [hereinafter An-Na'im, *Problems*] (discussing the exclusion of non-western participants and perspectives in the early development of the international human rights regime); HILARY CHARLESWORTH & CHRISTINE CHINKIN, THE BOUNDARIES OF INTERNATIONAL LAW: A FEMINIST ANALYSIS 36-37, 174-79 (2000) (identifying a "Southern" critique of the "Western origins, orientation and cultural bias" of the international legal order and noting that both the U.N. membership and its bureaucracy are dominated by men).

human rights law has attracted legitimate criticisms from those who were left out of the process of developing the international standards that they are now expected to embrace. Noted human rights scholar Abdullahi A. An-Na'im offers two main critiques: first, that the normative development of international human rights law has been limited by the dominant influence of the "cultures" that developed it; and second, that the real and perceived absence of cultural legitimacy at the domestic level has hindered implementation of international human rights law.[33] These are, of course, related points because the more international human rights law reflects accepted domestic cultural values, the greater the likelihood of successful implementation of those laws at the domestic level.[34]

An-Na'im has considered the importance of domestic cultural legitimacy for the successful implementation of international human rights standards primarily in areas of perceived conflict between human rights and Islam.[35] He

---

33. *See* An-Na'im, *State Responsibility*, *supra* note 12, at 171 (noting that the argument against universal cultural legitimacy of human rights on the ground that "the basic conception and major principles . . . emerged from western philosophical and political developments" may be true in light of factors such as "the nature and context of the drafting process, the limitations of studies purporting to cover a variety of cultural perspectives on the subject and the quality of representation of non-western points of view"); *see also* An-Na'im, *Problems*, *supra* note 32, at 331 (noting that discrepancies between theory and practice in the human rights arena results from ineffective enforcement of procedures).

34. Others have also considered the role of cultural legitimacy as it relates to increasing compliance with international law. *See* Koh, *supra* note 12, at 2600-02. Koh identifies the process as Transnational Legal Process (TLP), which "promotes the interaction, interpretation, and internalization of international legal norms." *Id.* at 2603. Koh distinguishes social, political, and legal internalization of norms, including explicit and implicit judicial internalization. *Id.* at 2656-57. Ryan Goodman and Derek Jinks also discuss coercion, persuasion, and acculturation as methods for ensuring compliance with international human rights law. Ryan Goodman & Derek Jinks, *How to Influence States: Socialization and International Human Rights Law*, 54 DUKE L.J. 621, 623 (2004). But others are more reluctant to endorse "internalization" of international norms. *See* Posner, *supra* note 8, at 34-36 (criticizing TLP on five grounds: "international law is not always good"; "judges are not always going to act the way we want them to"; "judges just don't know much about foreign policy, foreign countries, and foreign law"; "courts are, by design, passive and reactive"; and "TLP is undemocratic"). These themes are echoed in some of the dissenting opinions discussed *infra*.

35. In his article, *State Responsibility Under International Human Rights Law to Change Religious and Customary Laws*, An-Na'im uses his proposed framework to discuss changing or reinterpreting Islamic religious laws to better advance the human rights of women. An-Na'im, *State Responsibility*, *supra* note 12, at 181-84; *see also* Abdullahi A. An-Na'im, *Islam and Human Rights: Beyond the Universality Debate*, 94 ASIL PROC. 95, 99-100 (2000) (discussing religion and human rights in Mauritius); Abdullahi A. An-Na'im, *The Contingent Universality of Human Rights: The Case of Freedom of Expression in African and Islamic Contexts*, 11 EMORY INT'L L. REV. 29, 54-64 (1997) (discussing freedom of expression in Kenya and Sudan); Abdullahi A. An-Na'im, *The Rights of Women and International Law in the Muslim Context*, 9 WHITTIER L. REV. 491, 491-97 (1987) [hereinafter An-Na'im, *Rights of Women*] (discussing the rights of women under Shari'a law).

proposes a simple conceptual framework to evaluate and refine international human rights law and theory, both in terms of its future development and the success of efforts to implement it at the national level.[36] Under this framework, internal discourse plays a central role in developing the needed cultural legitimacy for human rights norms in the domestic arena, and cross-cultural dialogue offers a supporting role to external views in that process.[37]

This Article uses Professor An-Na'im's "internal discourse – cross-cultural dialogue" theoretical framework to describe and evaluate the current debate in the United States over the use of international human rights standards domestically.[38] Part I of this Article describes An-Na'im's model for evaluating and advancing the cultural legitimacy of human rights in domestic contexts. It briefly addresses the critique that current international human rights norms reflect historical and existing power relationships in the international community. But it focuses on the interplay of internal discourse and cross-cultural dialogue in the domestic implementation of those norms. Part II scrutinizes the current debate about the role of international human rights law in United States courts through this internal discourse – cross-cultural dialogue structure as part of the process of increasing the cultural legitimacy of human rights in the United States. This Part considers the recent Supreme Court decisions in *Atkins*, *Simmons*, and *Lawrence* that implicate both domestic constitutional law and, arguably, international human rights law: the prohibition of the death penalty for the developmentally disabled and for juveniles[39] and the affirmation of the freedom of individuals to engage in private sexual behavior.[40] This Part also briefly addresses the implications of

---

36. An-Na'im, *State Responsibility, supra* note 12, at 167-69; *see also* An-Na'im, *Problems*, *supra* note 32, at 331 (discussing discrepancies between law and practice in the human right arena, despite the existence of elaborate standards). He urges testing and use of his framework in other domestic contexts, as well. An-Na'im, *State Responsibility*, *supra* note 12, at 184-85.

37. An-Na'im, *State Responsibility, supra* note 12, at 174.

38. An-Na'im himself has been involved in examining the right to freedom of religion in the United States and is beginning to consider questions of cultural legitimacy of human rights in the United States. *See* An-Na'im, *Round Table Discussion, supra* note 14, at 975 (presenting the proceedings of a conference on religious liberty in the United States and including remarks by An-Na'im). This Article does not focus on the issue of religious rights but will draw upon some of the general insights offered by conference participants regarding United States culture and the domestic approach to human rights.

39. Part II.A. considers the recent Supreme Court decisions in *Atkins v. Virginia*, 536 U.S. 304 (2002) (prohibiting the execution of the mentally disabled) and *Roper v. Simmons*, 543 U.S. 551 (2005) (prohibiting the execution of juveniles). The death penalty raises domestic constitutional concerns under the Eighth Amendment's prohibition of "cruel and unusual punishment" and international human rights concerns under conventional and customary protection of the right to life and the prohibition of cruel, inhuman, or degrading treatment or punishment.

40. Part II.B. considers the recent Supreme Court decision in *Lawrence v. Texas*, 539 U.S. 558 (2003) (finding a Texas statute criminalizing certain sexual behavior between same-sex

the recent changes to the Supreme Court membership.[41] The Article concludes that both the Supreme Court's consideration of international and comparative law and the resulting controversy are appropriate aspects of the ongoing development of domestic law in accordance with international standards.

I. THE DEVELOPMENT OF DOMESTIC CULTURAL LEGITIMACY FOR INTERNATIONAL HUMAN RIGHTS NORMS AND LAWS

Understanding the process, and often the struggle, to develop cultural legitimacy for human rights standards at the domestic level requires an understanding of the recurring critiques of international human rights law, notably, that it is not truly "universal" and does not reflect widely shared values.[42] It is within this context that the framework developed by An'Na-im, with its emphasis on internal discourse and the supplementary role of cross-cultural dialogue,[43] takes on greater explanatory value and practical significance.

A. Is "International" the Same as "Universal"? An Overarching Critique of the Legitimacy of International Human Rights.

The United States was instrumental in the development of the modern international human rights movement.[44] To much of the world, the United

---

partners unconstitutional). The treatment of same-sex intimate relationships and conduct raises domestic constitutional concerns under the Fourteenth Amendment's protection of "liberty" interests and international human rights concerns under conventional and customary protection of the right to privacy and freedom from discrimination.

41. Part II.C. discusses the confirmation hearings of new Chief Justice John Roberts and new Justice Samuel Alito and the views they expressed regarding the use of international and foreign law in Supreme Court decisions.

42. *See infra* note 45 and accompanying text.

43. An-Na'im, *State Responsibility*, *supra* note 12, at 174.

44. Former United States President Franklin D. Roosevelt coined the name "United Nations," which was first used in the "Declaration by United Nations" of January 1, 1942. United Nations, History of the United Nations, http://www.un.org/aboutun/unhistory (last visited Jan. 12, 2007). "In 1945, representatives of 50 countries met in San Francisco at the United Nations Conference on International Organization to draw up the United Nations Charter." *Id.* Former United States President Harry S. Truman stated:

> My conviction that we had to have a world organization was so deep that I felt that no event, no matter how sad and unfortunate, should interfere with the drafting of the Charter of the United Nations. I felt that there was nothing I could do which would be more fitting to the memory of President Roosevelt than to go ahead with the conference.

Former President Harry S. Truman, Address at the Opera House 273 (June 24, 1955), www.un.org/depts/dhl/anniversary/stsg6j.pdf. "There have been women who have clearly left their mark on the history of the United Nations. Above all, there was Eleanor Roosevelt, who is linked to the Universal Declaration of Human Rights." Akmaral Arystanbekova, *Diplomacy: Too Important to be Left to Men?*, 39 UN CHRON. 62, 62 (Sept.-Nov. 2002), *available at*

Nations and the current international human rights legal regime profoundly reflect the values and priorities of the United States and other "western" democracies.[45] International and domestic human rights advocates often struggle with incorporating these "universal" values, which are frequently viewed as not universal at all, into particular domestic contexts. As a starting point, An-Na'im affirms the general importance of international human rights; yet he acknowledges the criticisms regarding the cultural bias inherent in the current system.[46] This cultural bias comprises both procedural and substantive strands: first, the failure to include other ("non-western") perspectives in the process of developing the initial human rights standards and system of

---

http://www.un.org/Pubs/chronicle/2002/issue3/0302p62_first_person.html. Mrs. Roosevelt also served as Chairperson of the United Nations Commission on Human Rights from 1946 through 1951. Chairpersons of the United Nations Commission on Human Rights, http://www.ohchr.org/english/bodies/chr/docs/FormerChairpersons.doc (last visited Jan. 12, 2007); *see also* An-Na'im, *Problems*, *supra* note 32, at 348-53 (discussing the exclusion of non-western participants and perspectives in the early era of the United Nations).

45. One critical perspective on human rights is that "human rights as propounded in the west are founded on individualism and therefore have no relevance to Asia which is based on the primacy of the community." Yash Ghai, *The Asian Perspective on Human Rights* (1993), www.hrsolidarity.net/mainfile.php/1993vol05no03/2061/ (1993). "Some Asian' [sic] governments consider that the western pressure on them for an improvement in human rights is connected with the project of western global hegemony," which is to be achieved "partly through the universalisation of western values and aspirations, and partly through the disorientation of Asian state and political systems." *Id.* The Chinese government has stated that "despite its international aspect, the issue of human rights falls by and large within the sovereignty of each state." *Id.* The Bangkok Governmental Declaration recognizes "that while human rights are universal in nature, they must be considered in the context of a dynamic and evolving process of international norm-setting, bearing in mind the significance of national and regional peculiarities and various historical, cultural and religious backgrounds." *Id.*; *see also* Makau Mutua, *Savages, Victims, and Saviors: The Metaphor of Human Rights*, 42 HARV. INT'L L.J. 201, 210 (2001) ("[H]uman rights, and the relentless campaign to universalize them, present a historical continuum in an unbroken chain of Western conceptual and cultural dominance over the past several centuries.").

46. An-Na'im, *State Responsibility*, *supra* note 12, at 172 (noting that there is already "significant consensus" on international human rights and that they do provide some "level of protection"). An-Na'im continues:

> It is neither possible, nor desirable in my view, for an international system of human rights standards to be culturally neutral. However, the claim of such an international system to universal cultural legitimacy can only be based on a moral and political "overlapping consensus" among the major cultural traditions of the world. In order to engage all cultural traditions in the process of promoting and sustaining such global consensus, the relationship between local culture and international human rights standards should be perceived as a genuinely reciprocal global collaborative effort.

Id. at 173.

protection;[47] and second, the "western" philosophical and theoretical origins of the current standards and system.[48] An-Na'im explains:

> Most African and Asian countries did not participate in the formulation of the Universal Declaration of Human Rights because, as victims of colonization, they were not members of the United Nations. When they did participate in the formulation of subsequent instruments, they did so on the basis of an established framework and philosophical assumptions adopted in their absence. For example, the pre-existing framework and assumptions favored individual civil and political rights over collective solidarity rights, such as a right to development, an outcome which remains problematic today. Some authors have gone so far as to argue that inherent differences exist between the Western notion of human rights as reflected in the international instruments and non-Western notions of human dignity. In the Muslim world, for instance, there are obvious conflicts between Shari'a and certain human rights, especially of women and non-Muslims.[49]

Many have questioned the notion of an exclusive "western" provenance to the ideas and values underpinning human rights standards.[50] But there is validity to critiquing the process's lack of inclusiveness, which involved a limited number of actors, most of whom were operating from a western perspective.

The United States presents different challenges to the cultural legitimacy of international human rights norms, in light of its different history in the international community. The United States has a history of inclusion, and even dominance, rather than exclusion. In its foreign policy, the United States implicitly and often explicitly expects other nations to conform their domestic practices to international law, including international human rights law.[51]

---

47. An-Na'im, *Problems*, *supra* note 32, at 346-53.

48. *Id.* at 346; An-Na'im, *State Responsibility*, *supra* note 12, at 171.

49. Abdullahi Ahmed An-Na'im, *Human Rights in the Muslim World*, *in* THE PHILOSOPHY OF HUMAN RIGHTS 315, 317 (Patrick Hayden ed., 2001).

50. *See* Catherine Powell, *Introduction: Locating Culture, Identity, and Human Rights*, 30 COLUM. HUM. RTS. L. REV. 201, 204-05 (1999) (discussing human rights principles as they relate to the traditions and philosophies of diverse non-western cultures (citing PAUL GORDON LAUREN, THE EVOLUTION OF INTERNATIONAL HUMAN RIGHTS: VISIONS SEEN 9-11 (1998))); *see also* Amartya Sen, *Human Rights and Asian Values*, THE NEW REPUBLIC, July 14 & 21, 1997, at 33, 40 ("Our ideas of political and personal rights have taken their particular form relatively recently, and it is hard to see them as 'traditional' commitments of Western cultures. . . . [A]ntecedents can be found plentifully in Asian cultures as well as Western cultures.").

51. For example, "no [foreign] assistance may be provided . . . to the government of any country which engages in a consistent pattern of gross violations of internationally recognized human rights, including torture or cruel, inhuman, or degrading treatment or punishment." 22 U.S.C. § 2151n(a) (2000). In addition, "none of the funds made available to carry out this chapter, and none of the local currencies generated under this chapter, shall be used to provide training or advice, or provide any financial support, for police, prisons, or other law enforcement forces for any foreign government or any program." 22 U.S.C. § 2420(a). The

United States civic, religious, and other non-governmental organizations do the same.[52] Nonetheless, the United States shares with many other nations the phenomenon of resisting outside views on how its domestic laws should operate.[53] In contrast to the Islamic nations of Africa and the Middle East, the United States played a significant role in the creation of the United Nations system and the International Bill of Rights; therefore, both should reflect values derived from and compatible with U.S. culture.[54] But many in the United States perceive that this is not the case. Thus, the acceptance of international

---

Secretary or the Administrator "shall not enter into any agreement under this chapter to provide agricultural commodities, or to finance the sale of agricultural commodities, to the government of any country determined by the President to engage in a consistent pattern of gross violations of internationally recognized human rights." 7 U.S.C. § 1733(j)(1). The United States has also applied human rights to grounds of inadmissibility. "Any alien who, while serving as a foreign government official, was responsible for or directly carried out, at any time, particularly severe violations of religious freedom, as defined in section 3 of the International Religious Freedom Act of 1998 (22 U.S.C. 6402), is inadmissible." Immigration and Nationality Act, Pub. L. No. 108-458, § 5502, 118 Stat. 3740, 3741 (2004).

52. Regarding Nigeria's adoption of the Shari'a, Human Rights Watch has stated, "Whatever personal beliefs may prevail in different social and religious circles in Nigeria, the Nigerian government—both at federal and state level—remains bound by international obligations and conventions." Human Rights Watch, *Failure to Conform to International Human Rights Standards* (Sept. 2004), www.hrw.org/reports/2004/nigeria0904/15.htm. "The UDHR is widely recognized as customary international law. It is a basic yardstick to measure any country's human rights performance. Unfortunately, Cuba does not measure up." Human Rights Watch, *Cuba's Repressive Machinery* (1999), http://www.hrw.org/reports/1999/cuba/Cuba996-01.htm. "Cuban courts continue to try and imprison human rights activists, independent journalists, economists, doctors, and others for the peaceful expression of their views, subjecting them to the Cuban prison system's extremely poor conditions." *Id.* "While Cuba's domestic legislation includes broad statements of fundamental rights, other provisions grant the state extraordinary authority to penalize individuals who attempt to enjoy their rights to free expression, opinion, press, association, and assembly." *Id.* It has been recommended that "[t]he Cuban government should cease all prosecutions based on the individual's exercise of fundamental rights to free expression, association, and movement," and that "[t]he Cuban government should reform its Criminal Code, repealing or narrowing the definition of all crimes that are in violation of established international human rights norms and practices." *Id.*

53. *See supra* notes 4, 15.

54. This should particularly be true with regard to civil and political rights. For example, the Declaration of Independence famously states: "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." THE DECLARATION OF INDEPENDENCE para. 2 (U.S. 1776). The Universal Declaration of Human Rights echoes this language in its Preamble ("Whereas recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world[.]"), in Article 1 ("All human beings are born free and equal in dignity and rights."), and Article 3 ("Everyone has the right to life, liberty and the security of person."), and in other provisions. UDHR, *supra* note 30, at 71-72.

human rights norms relies not solely on the nation's involvement in the historic and ongoing development of those norms at the international level, but also on the domestic process of acceptance and implementation.

B. Grounding the "Universal" in a Specific Cultural Context: Process and Bottom-Line Results.

Ultimately, the current international system must operate as a starting point for any analysis or proposed change. The existing human rights regime has many strengths, including the simple facts of its existence and endurance as a legal and political framework.[55] After more than sixty years of history, most participants and observers agree that certain shared fundamental values do exist, such as the prohibitions of genocide, slavery, torture, and systematic discrimination.[56] Now, the focus of critical reform efforts should be on increasing the legitimacy, effectiveness, and inclusiveness of the existing human rights legal regime:

> Since we already have an international system of human rights law and institutions, the process should seek to legitimize and anchor the norms of this established system within, and between, the various cultural traditions of the

---

55. One of the basic purposes of the United Nations since its founding in 1945 has included the protection of internationally recognized human rights. U.N. Charter art. 1, para. 3. In 1946, the United Nations General Assembly established a Commission on Human Rights, and in 1948 it adopted the Universal Declaration of Human Rights. The United Nations and its member states are currently engaged in efforts to reform the United Nation's existing human rights mechanisms by creating a new Human Rights Council. *See* Warren Hoge, *With Its Human Rights Oversight Under Fire, U.N. Submits a Plan for a Strengthened Agency*, N.Y. TIMES, Feb. 3, 2006, at A6, *available at* 2006 WLNR 1901893.

56. These are considered *jus cogens* norms, part of binding, customary and conventional international law. *See* RESTATEMENT (THIRD) OF THE FOREIGN RELATIONS LAW OF THE UNITED STATES § 702 (1987). *See generally* Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment, G.A. Res. 39/46, at 197, U.N. GAOR, 39th Sess., U.N. Doc. A/39/51 (Dec. 10, 1984) (establishing that torture is never justified under any circumstance); Convention on the Elimination of All Forms of Discrimination Against Women, G.A. Res. 34/180, at 193, U.N. GAOR, 34th Sess., U.N. Doc. A/34/46 (Dec. 18, 1979) [hereinafter CEDAW] (promoting equal rights and condemning any distinction based on sex); International Convention on the Elimination of All Forms of Racial Discrimination, G.A. Res. 2106A, at 47, U.N. GAOR, 20th Sess., U.N. Doc. A/6014 (Dec. 21, 1965) (recognizing that racial discrimination is repugnant to human society); Convention on the Prevention and Punishment of the Crime of Genocide, G.A. Res. 260A, at 174, U.N. GAOR, 3d Sess., U.N. Doc. A/810 (Dec. 9, 1948) (confirming that genocide is a crime under international law); International Convention to Suppress the Slave Trade and Slavery, Sept. 25, 1926, 46 Stat. 2183, 60 L.N.T.S. 253, *available at* http://www1.umn.edu/humanrts/instree/f1sc.htm (affirming the intent to abolish slavery in all its forms throughout the world). An-Na'im uses the shared fundamental principles of nondiscrimination and legal equality in his work on increasing human rights protections for women and non-Muslims under Shari'a law. *See, e.g.*, An-Na'im, *State Responsibility*, *supra* note 12, at 181-82; An-Na'im, *Rights of Women*, *supra* note 35, at 502.

> world. In other words, the norms of the international system should be validated in terms of the values and institutions of each culture, and also in terms of shared or similar values and institutions of all cultures. This can be achieved, I suggest, through what I call "internal discourse" within the framework of each culture, and "cross-cultural dialogue" among the various cultural traditions of the world.[57]

"Internal discourse," defined as the discussion and debate on the application and implementation of human rights at the domestic level, is important in several respects. Internal discourse must serve as the primary mechanism for advancing the cultural legitimacy of human rights norms.[58] Human rights are presumed to be protected initially and primarily at the domestic level.[59] Additionally, internal discourse is important as a means to educate and socialize, which aids in developing both consensus and credibility for the standards adopted.[60] Internal discourse serves as a global equalizer of sorts, a manifestation of the reciprocal nature of the process. A nation with a visible internal process draws attention to its treatment of human rights issues, and also serves as a model for other nations.[61] At the same time, this visible

---

57. An-Na'im, *State Responsibility*, *supra* note 12, at 174. An-Na'im does not necessarily endorse the desirability of all existing human rights norms but instead is interested in using them as a starting point for further development.

58. *See id.* at 169 (given "the nature of international law in general, and its dependence on largely voluntary compliance and cooperation of sovereign states in the field of human rights in particular," it is necessary to seek greater consensus in the domestic cultural context). Accordingly, decisions about how to challenge particular domestic laws should be left to the process of internal discourse.

59. *See supra* note 24.

60. *See* An-Na'im, *State Responsibility*, *supra* note 12, at 174, 178-79. Addressing this issue in the context of reconciling international human rights with customary and religious laws in Muslim countries, An-Na'im explains:

> [I]t is clear that the only viable and acceptable way of changing religious and customary laws is by transforming popular beliefs and attitudes, and thereby changing common practice. This can be done through a comprehensive and intensive program of formal and informal education, supported by social services and other administrative measures, in order to change people's attitudes about the necessity or desirability of continuing a particular religious or customary practice. To achieve its objective, the program must not only discredit the religious or customary law or practice in question, but also provide a viable and legitimate alternative view of the matter.

Id. at 178.

61. An-Na'im provides two main rationales for the importance of internal discourse:

> First, internal validation is necessary in all cultural traditions for one aspect or another of the present international human rights system. . . . Second, for such discourse within one culture to be viable and effective, its participants should be able to point to similar discourse which is going on in the context of other cultures.

Id. at 174.

internal process underscores the notion that no nation is immune from having to adjust its national system to comport with international obligations.[62]

The mechanism of cross-cultural dialogue functions as a significant counterpart and supplement to internal discourse.[63] External actors may directly encourage dialogue in numerous ways, such as "international action to protect the freedoms of speech and assembly of internal actors . . . and assistance in developing and implementing campaign strategies."[64] Cross-cultural dialogue may also involve strategic use of external resources, such as the laws and practices of different nations or "the exchange of insights and experiences about the concept of the particular human right and the sociopolitical context of its implementation" by internal actors.[65] In addition, reciprocity plays a role in increasing acceptance of cross-cultural input and in establishing basic thresholds for compliance with human rights standards.[66] But cross-cultural dialogue must function primarily as a supplement to internal discourse. There is always a risk that if external actors take an overly active role in the internal process of reinterpreting cultural norms, their involvement may be counter-productive and may even alienate those generally supportive of human rights norms.[67] Again, this is particularly true in areas that are suspicious of "western" influence based on a history of colonialism and religious conflict.[68]

---

62. *See id.* (explaining that internal validation "might be necessary for civil and political rights in one culture, economic and social rights in another, the rights of women or minorities in a third, and so forth").

63. *See id.* An-Na'im offers two main rationales regarding the importance of cross-cultural dialogue:

> First, from a methodological point of view, all participants in their respective internal discourses can draw on each other's experiences and achievements. Second, cross-cultural dialogue will enhance understanding of, and commitment to, the values and norms of human dignity shared by all human cultures, thereby providing a common moral and political foundation for international human rights standards.

Id.

64. *Id.* at 179. For example, the use of letter-writing or "urgent action" campaigns by Amnesty International and other non-governmental organizations are often directed at protecting the right to freedom of expression. *See, e.g.*, Amnesty International, *Act Now for Human Rights*, http://www.amnesty.org/actnow/ (last visited Jan. 12, 2007) (including appeals to protect "human rights defenders").

65. An-Na'im, *State Responsibility*, *supra* note 12, at 179.

66. *See* An-Na'im, *Problems*, *supra* note 32, at 345. For instance, in his approach to reconstructing the relationship between Islam and human rights, An-Na'im focuses on particular examples, such as the treatment of women and non-believers. *See id.* In arguing for a reinterpretation of existing Islamic law, he suggests incorporation of the principle of reciprocity as one way of reconciling the nondiscrimination principles of human rights with current interpretations of Shari'a law. *Id.*

67. An-Na'im, *State Responsibility*, *supra* note 12, at 179-80.

68. *See id.* at 184; *see also* Deborah M. Weissman, *The Human Rights Dilemma: Rethinking the Humanitarian Project*, 35 COLUM. HUM. RTS. L. REV. 259, 291 (2004)

Although An-Na'im employs the internal discourse – cross-cultural dialogue construct to suggest means of reconciling international human rights standards with Islamic law in particular national contexts, this theoretical framework has broader application, as well. In this era of globalization, all nations face a similar challenge—to a greater or lesser degree—of reconciling international human rights standards with their existing national legal system, which incorporates both the formal legal structure and the informal influence of culture, religion, and custom.[69] Resistance to human rights standards generally, or more commonly to particular rights, is rooted in this informal system of norms and values.[70] Moreover, the resistance to international law at the domestic level rarely centers on the content of the norm itself, but rather on the interpretation and specific implementation of the norm in the particular domestic situation.[71] Appreciation of this strategic interplay of internal

---

("Formerly subjugated people's suspicions of human rights values emanating from colonial powers must be viewed as part of the legacy of colonialism. The logic of these suspicions is easy to discern, as people denied autonomy seek to establish cultural self-determination. Misgivings about a human rights agenda originating from former colonizers is not unreasonable.").

69. This expansive obligation is often reflected in general treaty language. *See, e.g.*, ICCPR, *supra* note 10, at art. 2; *see also* CEDAW, *supra* note 56, at art. 3 ("States Parties shall take in all fields, in particular in the political, social, economic and cultural fields, all appropriate measures, including legislation, to ensure the full development and advancement of women . . . .").

70. An-Na'im's definition of culture broadly includes religion and custom. An-Na'im, *Problems*, *supra* note 32, at 335-36. He notes:

> The prime feature underlying cultural legitimacy is the authority and reverence derived from internal validity. A culturally legitimate norm or value is respected and observed by the members of the particular culture, presumably because it is assumed to bring satisfaction to those members. Because there may be conflicts and tensions between various competing conceptions of individual and collective satisfaction, there is constant change and adjustment of the norms or values in any culture which are accorded respect and observance.

Id. at 336. Since this informal system of norms and values is particularly powerful in the context of discrimination against women, the obligations of parties to CEDAW are especially broad. The CEDAW Resolution states:

> States Parties shall take all appropriate measures:
>
> (a) To modify the social and cultural patterns of conduct of men and women, with a view to achieving the elimination of prejudices and customary and all other practices which are based on the idea of the inferiority or the superiority of either of the sexes or on stereotyped roles for men and women.

CEDAW, *supra* note 56, at art. 5(a). The Resolution also states that CEDAW will "take all appropriate measures, including legislation, to modify or abolish existing laws, regulations, customs and practices which constitute discrimination against women." *Id.* at art. 2(f).

71. For example, few nations or activists currently would contest a general prohibition on discrimination on the basis of race or gender. But many would contest the meaning of that general prohibition in their particular domestic contexts. An-Na'im discusses the interpretation of gender-based discrimination under Islamic law. *See* An-Na'im, *State Responsibility*, *supra*

discourse and cross-cultural dialogue[72] offers a simple methodology for approaching the complexities of implementing human rights standards in the wide variety of domestic contexts in today's global community and for ensuring the cultural legitimacy—and ultimate effectiveness—of those standards.

II. A HUMAN RIGHTS CULTURE IN THE UNITED STATES SUPREME COURT

The role of the executive branch in international law is familiar: treaty-making, foreign policy development, and diplomacy.[73] The role of the legislative branch is familiar, as well: providing advice and consent to treaties, and adopting implementing legislation.[74] But less attention has been given to the important role of domestic courts in implementing the nation's international obligations, particularly its international human rights law obligations.[75] Domestic courts may directly apply international standards,[76] but given the general nature of the standards at the international level, domestic courts also play a role by interpreting those standards in their domestic context.[77] The

---

note 12, at 181. In the United States, there are ongoing debates about the appropriateness of affirmative action programs. *See* Gratz v. Bollinger, 539 U.S. 244, 268-71 (2003); Grutter v. Bollinger, 539 U.S. 306, 326-27 (2003). Note that Justice Ginsberg cites to international human rights law in her concurrence in *Grutter*. *Id.* at 344 (Ginsburg, J., concurring).

72. *See* An-Na'im, *State Responsibility*, *supra* note 12, at 174 (stating that "the combination of the processes of internal discourse and cross-cultural dialogue will, it is hoped, deepen and broaden universal cultural consensus on the concept and normative content of international human rights").

73. RESTATEMENT (THIRD) OF THE FOREIGN RELATIONS LAW OF THE UNITED STATES § 303 (describing the President's role in authorizing or approving international agreements).

74. *Id.* (describing Congress's role in authorizing or approving international agreements).

75. *See id.* § 115 n.3 (explaining that "courts will give effect to international law 'where there is no treaty, and no controlling executive or legislative act or judicial decision,' and 'in the absence of any treaty or other public act of their own government in relation to the matter'" (quoting The Paquete Habana, 175 U.S. 677, 700, 708 (1900)). The high courts of a few nations have been the exception; for example, the post-apartheid South African Supreme Court and the Canadian Supreme Court frequently cite to materials from other nations to support their holdings. *See* Sarah K. Harding, *Comparative Reasoning and Judicial Review*, 28 YALE J. INT'L L. 409, 416-17 (2003) (providing a case study of the Supreme Court of Canada); Cody Moon, Note, *Comparative Constitutional Analysis: Should the United States Supreme Court Join the Dialogue?*, 12 WASH. U. J.L. & POL'Y 229, 232-39 (2003) (discussing the high courts of Canada, South Africa and Australia). Additionally, other high courts have drawn notice for high profile cases such as the British, Spanish and Chilean courts considering the Pinochet prosecution. *See* Amnesty International, *The Case of Augusto Pinochet: Timeline*, http://news.amnesty.org/pages/pinochet_timeline (last visited Jan. 12, 2007) (highlighting the involvement of the United Kingdom and Chilean courts in the *Pinochet* case).

76. In United States courts, this happens when treaty provisions are self-executing or when the court applies customary international law. RESTATEMENT (THIRD) OF FOREIGN RELATIONS LAW OF THE UNITED STATES § 111.

77. This is specifically accounted for in the doctrine of "margin of appreciation" employed by the European Court of Human Rights and other international human rights bodies:

Supreme Court has had opportunities to ensure the direct implementation of international human rights standards domestically (and in turn, to play a role in developing international jurisprudence on human rights), but the Court has been cautious about doing so thus far.[78]

Nevertheless, the Court has increasingly played another perhaps more important role in advancing the cultural legitimacy of particular human rights standards. In the Supreme Court, Congress, and the public life, the current debates over the use of international law are truly about cultural legitimacy of the international human rights standards, rather than domestic implementation.[79] What generates the passion is not whether a particular treaty or norm is self-executing, but whether the treaty or norm represents the values

---

> Although the Commission and Court [now Court] invoke the principle of strict interpretation and thus the favourable balancing of individual rights against state interests, they in fact leave a certain amount of discretion for the states to decide whether a given course of action is compatible with Convention requirements. This state discretion is referred to as the "margin of appreciation."

DONNA GOMIEN, DAVID HARRIS & LEO ZWAAK, LAW AND PRACTICE OF THE EUROPEAN CONVENTION ON HUMAN RIGHTS AND THE EUROPEAN SOCIAL CHARTER 215 (Council of Europe Publishing 1996). The doctrine reflects the view that "state authorities are in principle in a better position than the international judge" to assess the balance of rights in particular domestic contexts. *Id.* (quoting Handyside v. United Kingdom, 24 Eur. Ct. H.R. (ser. A) at 22 (1976)). Moreover, persuasive domestic jurisprudence finds its way back into the decisions of international bodies. For example, the European Court of Human Rights provides a list of its cases where it has drawn upon comparative law in its reasoning. *See* European Court of Human Rights, Selective Comparative Law Case List, http://www.echr.coe.int/ECHR/EN/Header/Case-Law/Case-law+information/Selective+comparative+law+case+list/ (last visited Jan. 12, 2007).

78. *See* Sosa v. Alvarez-Machain, 542 U.S. 692, 720-21 (2004) (adopting a narrow understanding of "the law of nations" in determining the scope of the Alien Tort Statute).

79. The Supreme Court has a mixed history regarding the protection of individual rights. There are cases that pre-date the modern international human rights movement that still address important human rights issues including discrimination. *See generally* Korematsu v. United States, 323 U.S. 214 (1944) (analyzing the legality of a civilian exclusion order that curtailed the rights of a particular ethnic group); Plessy v. Ferguson, 163 U.S. 537 (1896) (finding that a statute that distinguished between races was constitutional); Bradwell v. State, 83 U.S. 130 (1872) (affirming the denial of a woman's law practice application); Dred Scott v. Sandford, 60 U.S. 393 (1856) (dismissing an action brought by a slave because he was not considered a citizen). More recently, however, the Court has adopted a new mentality that progressively defines and even extends individual rights protection. *See generally* Roe v. Wade, 410 U.S. 113 (1973) (finding that abortion is within the scope of personal liberty guaranteed by the Constitution); Brown v. Bd. of Educ., 347 U.S. 483 (1954) (finding that segregation was unconstitutional). But within the Court, much debate remains regarding the issue of domestic implementation and application of international law. For example, in another recent and controversial Supreme Court decision, the issue was not cultural legitimacy, but rather the complications of discerning and applying international human rights norms where there is statutory authorization to do so under the Alien Tort Statute. *Sosa*, 542 U.S. at 727-28. In that case, the Court did not focus on whether "international" norms made sense and helped us understand our own law, but on whether the norms could be directly applied. *Id.*

of the United States or the imposed values of outsiders.[80] The Court's decisions in *Atkins*, *Simmons*, and *Lawrence* embody and advance this ongoing internal discourse – cross-cultural dialogue process.

The United States Supreme Court's opinions address the criminal law issues of the death penalty and the criminalization of particular sexual conduct, but they also raise important constitutional and human rights questions about individual liberty, privacy, and dignity.[81] Both areas implicate values traditionally associated with the U.S. criminal justice system (deterrence, punishment, retribution, rehabilitation, and accountability) and strong religious and cultural values (vengeance, mercy, procreation, and the sanctity of intimate relationships).

Internal discourse regarding the imposition of the death penalty is active and enduring in the United States;[82] the treatment of same-sex relationships and sexual conduct is also the subject of wide-ranging domestic debate.[83] The

---

80. A common criticism of the use of international and foreign legal sources in the decisions has been that such use is "undemocratic," or anti-democratic. *See e.g.*, Atkins v. Virginia, 536 U.S. 304, 322 (2002) (Rehnquist, C.J., dissenting) (stating that the Court's use of foreign laws is "antithetical to considerations of federalism"). *But see* Darren Lenard Hutchinson, *The Majoritarian Difficulty: Affirmative Action, Sodomy, and Supreme Court Politics*, 23 LAW & INEQ. 1, 32, 42-44 (2005) (arguing that the view of the Supreme Court as a counter-majoritarian institution is flawed and examining the Court's decision in *Lawrence v. Texas* as an example of the Court's majoritarian nature).

81. This process of cultural legitimization also is demonstrated in other contexts, such as the Court's consideration of the "death row phenomenon," and affirmative action. *See* Gratz v. Bollinger, 539 U.S. 244, 275 (2003); Grutter v. Bollinger, 539 U.S. 306, 329 (2003); Foster v. Florida, 537 U.S. 990, 990 (2002); Knight v. Florida, 528 U.S. 990, 990 (1999).

82. There are numerous organizations that are active in opposition to the death penalty in the United States. For example, Amnesty International lists over twenty organizations in the United States. Amnesty International, *External Links: Death Penalty*, http://web.amnesty.org/links/bytheme?readform&restricttocategory=DEATH+PENALTY&count=30 (last visited Jan. 12, 2007); *see also* Death Penalty Information Center, http://www.deathpenaltyinfo.org/ (last visited Jan. 12, 2007) (providing information on issues concerning death penalty). Twelve states and the District of Columbia have no death penalty. Amnesty International USA, Abolish the Death Penalty, http://www.amnestyusa.org/abolish/states/index.html (last visited Jan. 12, 2007). Other organizations actively support the death penalty. ProDeathPenalty.com lists numerous organizations that support the death penalty and provides links to supportive or neutral sites. ProDeathPenalty.com, *Death Penalty Links*, http://www.prodeathpenalty.com/links2.htm (last visited Jan. 12, 2007). The Prosecutor's Office in Clark County, Indiana, also has a website that provides a statement supporting the death penalty and provides links to other pro-death penalty sites. Steven D. Stewart, *A Message from the Prosecuting Attorney*, http://www.clarkprosecutor.org/html/death/death.htm (last visited Jan. 12, 2007).

83. There are also many organizations active on issues related to same-sex relationships. The Human Rights Campaign is a prominent advocacy group that works for lesbian, gay, bisexual and transgender equal rights. *See* Human Rights Campaign, Working for Lesbian, Gay, Bisexual and Transgender Equal Rights, http://www.hrc.org/ (last visited Jan. 12, 2007). Others are active in opposition. *See, e.g.*, Focus on the Family ACTION, http://www.focusaction.org/

Supreme Court's decisions in these areas reflect internal discourse and ultimately contribute to it. The decisions have generated controversy largely due to their content and outcomes, but also due in part to their explicit references to foreign and international law in support of their outcomes. The cross-cultural dialogue that appears in the decisions has been used as merely a supplement, yet it has provoked public outcry. On the issue of the death penalty, U.S. law and policy appears out-of-step with the worldwide trend towards abolition;[84] on the issue of privacy and liberty in intimate conduct, the United States is arguably in the vanguard for protecting individual rights.[85]

A. The "evolving standards of decency that mark the progress of a maturing society"[86]: Lessons from Atkins and Simmons about the Human Right to Life.

From an international human rights law perspective, although the standards regarding the death penalty continue to evolve, the standards are relatively well-established. There is a general consensus internationally that the death penalty conflicts with the right to life and the right to security of the person; thus, the death penalty should be abolished.[87] If exceptions are allowed, they are quite

---

(last visited Jan. 12, 2007).

84. According to Amnesty International, 128 countries have abolished the death penalty by law or in practice. Of the remaining 69 countries, only a small number actually execute prisoners, and the global trend is towards abolition. Amnesty International, *Facts and Figures of the Death Penalty*, http://web.amnesty.org/pages/deathpenalty-facts-eng (last visited Jan. 12, 2007).

85. Several European countries provide some legal recognition of same-sex relationships including: Denmark, Norway, Sweden, Iceland, Finland, the Netherlands, Belgium, Spain, Germany, France, Luxembourg and Britain. *Gay Marriage Around the Globe*, BBC NEWS, Dec. 22, 2005, http://news.bbc.co.uk/1/hi/world/americas/4081999.stm. Other countries that provide either marriage or similar recognition include Canada, Argentina, New Zealand, and South Africa. *Id.* The Human Rights Campaign also lists Croatia, Hungary, Israel, Portugal, Slovenia, and Switzerland as providing some legal recognition or benefits to same-sex couples. *See* Human Rights Campaign, *Marriage/Relationship Recognition Laws: International*, http://hrc.org/Template.cfm?Section=Center&CONTENTID=26546&TEMPLATE=/TaggedPage/TaggedPageDisplay.cfm&TPLID=70 (last visited Jan. 12, 2007).

86. *See* Trop v. Dulles, 356 U.S. 86, 101 (1958) (plurality opinion), *quoted in* Roper v. Simmons, 543 U.S. 551, 561 (2005); Atkins v. Virginia, 536 U.S. 304, 311-12 (2002).

87. All major human rights treaties provide for the right to life, the right to security of the person, and the right to be free from cruel, inhuman or degrading treatment. *See* ICCPR, *supra* note 10, at art. 6, 7, 9 ("Every human being has the inherent right to life. . . . No one shall be subjected to torture or to cruel, inhuman or degrading treatment or punishment. . . . Everyone has the right to liberty and security of person."); African Charter, *supra* note 24, at art. 4, 5, 6 ("Every human being shall be entitled to respect for his life and the integrity of his person. . . . All forms of . . . cruel, inhuman or degrading punishment and treatment shall be prohibited. . . . Every individual shall have the right to liberty and to the security of his person."); American Convention on Human Rights: "Pact of San José, Costa Rica," art. 4, 5, 7, Apr. 8, 1970, 1144 U.N.T.S. 144, 145-46 [hereinafter ACHR] ("Every person has the right to have his life

narrow and framed within a context of moving toward eventual abolition.[88] This is certainly true for the nations that the United States views as sharing its history, traditions and values.[89] Moreover, most of the world favors abolishing the death penalty.[90] The international preference for abolishing the death penalty is magnified when examining the specific contexts of execution of the developmentally disabled or of juveniles.[91] Although many international

---

respected. . . . No one shall be subjected to torture or to cruel, inhuman, or degrading punishment or treatment. . . . Every person has the right to personal liberty and security."); Convention for the Protection of Human Rights and Fundamental Freedoms art. 2, 3, 5, Nov. 4, 1950, 213 U.N.T.S. 222, 224, 226 [hereinafter European Convention] ("Everyone's right to life shall be protected by law. . . . No one shall be subjected to torture or to inhuman or degrading treatment or punishment. . . . Everyone has the right to liberty and security of person.").

88. For example, after the general assertion in Article 6 that "[e]very human being has the inherent right to life," the ICCPR further provides:

> 2. In countries which have not abolished the death penalty, sentence of death may be imposed only for the most serious crimes in accordance with the law in force at the time of the commission of the crime. . . .
>
> . . . .
>
> 4. Anyone sentenced to death shall have the right to seek pardon or commutation of the sentence. Amnesty, pardon or commutation of the sentence of death may be granted in all cases.
>
> . . . .
>
> 6. Nothing in this [A]rticle shall be invoked to delay or to prevent the abolition of capital punishment by any State Party to the present Covenant.

ICCPR, *supra* note 10, at art. 6. The Second Optional Protocol to the ICCPR is specifically directed towards the abolition of the death penalty and provides, "Each State Party shall take all necessary measures to abolish the death penalty within its jurisdiction." Second Optional Protocol to the International Covenant on Civil and Political Rights art. 1 (entered into force on July 11, 1991).

89. For example, the European Convention sets out an explicit right to life in Article 2: "Everyone's right to life shall be protected by law. No one shall be deprived of his life intentionally save in the execution of a sentence of a court following his conviction of a crime for which this penalty is provided by law." European Convention, *supra* note 87, at art. 2. In addition, the Convention has established a specific protocol regarding the death penalty. *See* Protocol No. 6 to the Convention for the Protection of Human Rights and Fundamental Freedoms Concerning the Abolition of the Death Penalty, C.E.T.S. No. 114 (ratified by all members of the Council of Europe except Russia).

90. *See supra* note 87. For an up-to-date report of which countries currently practice the death penalty and which have abolished or abandoned it, see Amnesty International, *Abolitionist and Retentionist Countries*, http://web.amnesty.org/pages/deathpenalty-countries-eng (last visited Jan. 12, 2007).

91. Several countries have recently raised the minimum age for the death penalty to 18 years old, including: Yemen, Zimbabwe, China and Pakistan. Amnesty International, *Stop Child Executions!*, http://web.amnesty.org/pages/deathpenalty-children-eng (last visited Jan. 12, 2007). Recent decisions by the Inter-American Commission on Human Rights have held that the prohibition of the execution of juvenile offenders constitutes a *jus cogens* norm. Domingues v. United States, Case 12.285, Inter-Am. C.H.R., Report No. 62/02, ¶¶ 84-85 (2002), *available at* www.cidh.org; Beazley v. United States, Case No. 12.412, Inter-Am. C.H.R., Report No.

human rights norms appear to lack consensus, either by treaty participation or by state practice, the norms regarding the human rights violations inherent in the death penalty are widely agreed upon. In fact, the United States is an anomaly regarding its death penalty doctrine.[92]

Nonetheless, internal discourse about the death penalty continues to occur, particularly in the United States.[93] In fact, internal discourse is implicit in the standards set forth in the Constitution; the Eight Amendment's prohibition of "cruel and unusual" punishment[94] suggests an ongoing evaluation of the punishment against other standards to ensure that it is "graduated and proportioned to [the] offense."[95] In that evaluation, the Supreme Court established a standard that considers "the evolving standards of decency that mark the progress of a maturing society."[96] Not only does the Court's standard contain an inherently comparative function, but the Court also firmly grounds its standard in "the duty of the government to respect the dignity of all persons."[97]

Although the United States has not abolished the death penalty,[98] the Supreme Court has recently limited its scope: in *Atkins v. Virginia*,[99] the Court

---

101/03, ¶¶ 47-50 (2003), *available at* www.cidh.org. The ICCPR, Article 6, also provides, "Sentence of death shall not be imposed for crimes committed by persons below eighteen years of age and shall not be carried out on pregnant women." ICCPR, *supra* note 10, at art. 6.

92. In the context of the juvenile death penalty, the Court in *Simmons* stated that "it is fair to say that the United States now stands alone in a world that has turned its face against the juvenile death penalty." Roper v. Simmons, 543 U.S. 551, 577 (2005). Although some may suggest that the United States may ignore international human rights law because the country already has a culture of compliance with individual rights, that argument is not available in this context.

93. The Supreme Court is, of course, not the only participant in this discourse. *See supra* note 82 (listing active organizations that advocate in favor of and against the death penalty). The United States' resistance to international pressure to abolish or limit the death penalty may reflect cultural values of independence and individual accountability, and cultural influences of fear, violence, and racism. Nonetheless, domestic and international advocates of human rights standards have been somewhat effective in changing the nation's views in this area, often by using the method An-Na'im articulated, which reinterprets values embraced by and in the United States to argue against the death penalty.

94. The Eighth Amendment provides: "Excessive bail shall not be required, nor excessive fines imposed, nor cruel and unusual punishments inflicted." U.S. CONST. amend. VIII.

95. *Simmons*, 543 U.S. at 560 (quoting Weems v. United States, 217 U.S. 349, 367 (1910)).

96. *Id.* at 561 (quoting Trop v. Dulles, 356 U.S. 86, 101 (1958) (plurality opinion)).

97. *Id.* at 560 ("By protecting even those convicted of heinous crimes, the Eighth Amendment reaffirms the duty of the government to respect the dignity of all persons.").

98. In 1972, the Supreme Court found that the death penalty violated the Eighth Amendment when its imposition was left to the sole discretion of juries; consequently, all states were required to rewrite their death penalty laws. Furman v. Georgia, 408 U.S. 238, 239-40 (1972). But in 1976, the Court allowed executions to resume in the United States. Gregg v. Georgia, 428 U.S. 153, 169 (1976).

99. 536 U.S. 304, 321 (2002).

prohibited the execution of the developmentally disabled, and in *Roper v. Simmons*,[100] the Court prohibited the execution of juveniles. In both *Atkins* and *Simmons*, the Court evaluated the historical debate about the death penalty in the United States and, noting the evolution of public sentiment, reversed earlier decisions.[101] The Court reviewed the "objective indicia of consensus" on the death penalty by looking to the actions of the individual states, professional associations, non-profit organizations, academia, and others with relevant information or perspective in assessing the "evolving standards of decency."[102] The Court's reliance on international materials in the *Atkins* and *Simmons* opinions created a greater controversy than the actual holdings themselves.[103] The Court looked beyond resources in the United States to the practices of other nations,[104] as well as the various perspectives within the international community.[105] In both *Atkins* and *Simmons*, the Court's analysis of the

---

100. 543 U.S. 551, 578 (2005).

101. *Id.* at 564-69; *Atkins*, 536 U.S. at 313-17.

102. *Simmons*, 543 U.S. at 563, 567. The majority of the states prohibited the juvenile death penalty at the time of the decision in *Simmons*. *Id.* at 568; *see also* HOWARD N. SNYDER & MELISSA SICKMUND, JUVENILE OFFENDERS AND VICTIMS: 1999 NATIONAL REPORT 89, 133 (National Center for Juvenile Justice 1999). *But see* Elizabeth S. Scott & Thomas Grisso, *The Evolution of Adolescence: A Developmental Perspective on Juvenile Justice Reform*, 88 J. CRIM. L. & CRIMINOLOGY 137, 148 (1997). The Court referred to numerous *amici curiae* in both *Atkins* and *Simmons*. For example, in *Atkins*, *amici* included the American Psychological Association, the United States Catholic Conference, and the European Union. *Atkins*, 536 U.S. at 316-17 n.21. In *Simmons*, *amici* included the European Union, President James Earl Carter, Jr., Former U.S. Diplomats, and the Human Rights Committee of the Bar of England and Wales. *Simmons*, 543 U.S. at 576.

103. The controversy also exists within the Court. Some Justices, notably Breyer, Kennedy and Ginsburg, support the use of international and comparative materials. See *supra* note 4; *infra* note 194. Other Justices are deeply opposed to the use of international and comparative materials, notably Scalia and Thomas. Justice Kennedy wrote the majority opinions in *Lawrence v. Texas* and *Roper v. Simmons*; Justice Stevens wrote the opinion of the Court in *Atkins v. Virginia*. Chief Justice Roberts has not yet had an occasion to address this issue in an opinion. But in his confirmation hearings, Chief Justice Roberts responded to questions about the use of international law in court decisions in the United States by stating: "I don't think it's a good approach... I'd accuse them of getting it wrong on that point, and I'd hope to sit down with them and debate it and reason about it." *Roberts Confirmation Hearing, supra* note 18, at 293 (statement of Judge John Roberts). He cited concerns arising under "democratic theory" and concerns regarding the problems of selectivity and lack of restraints on judicial discretion. *Id.* Justice Alito expressed similar, yet stronger, views regarding the same. *Alito Confirmation Hearing, supra* note 3; *see infra* Part II.C.

104. The Court's references to international views and practices did not begin in the *Simmons* and *Atkins* opinions. In *Thompson v. Oklahoma*, 487 U.S. 815 (1988), the Court looked to the views of "other nations that share our Anglo-American heritage, and by the leading members of the Western European community." *Id.* at 830.

105. There have been an increasing number of direct interventions by other States on this issue. Most recently, other countries have used the Vienna Convention on Consular Relations to intervene in the cases of foreign death row inmates in the United States. *See, e.g.*, Avena and

"evolving standards of decency"[106] draws heavily upon the internal discourse regarding the death penalty and selectively deploys cross-cultural dialogue to increase the cultural legitimacy of the decision.

In *Atkins*, the Supreme Court addressed the issue of the constitutionality of the death penalty for the developmentally disabled.[107] In 1989, the Court had examined this issue in *Penry v. Lynaugh* and concluded that the constitutional prohibition of cruel and unusual punishment did not preclude the execution of those who are developmentally disabled.[108] But fourteen years later, the Court reached the opposite conclusion in *Atkins*.[109] The main emphasis in *Atkins* was the changing nature of the internal discourse regarding the execution of those who are developmentally disabled.[110] The Court detailed the actions of state legislatures to prohibit the imposition of the death penalty for the developmentally disabled in the years after its decision in *Penry*.[111] The Court considered not only the number of states that had changed their laws, but also the "consistency of the direction of change" to determine that a national consensus had developed against the practice.[112] The Court's only reference to a cross-cultural influence was in a footnote addressing the views of the "world community."[113] But in the same note, the Court addressed the opinions of domestic professional and religious organizations, as well as polling data regarding the opinions of the American public.[114] The Court briefly noted that "within the world community, the imposition of the death penalty for crimes committed by mentally retarded offenders is overwhelmingly disapproved."[115]

Given the majority's limited inclusion of cross-cultural perspectives in *Atkins*, the depth in which the dissenting opinions addressed the matter is quite surprising. In his dissent, then Chief Justice Rehnquist (joined by Justices Scalia and Thomas), wrote separately to "call attention to the defects in the Court's decision to place weight on foreign laws, the views of professional and

---

Other Mexican Nationals (Mexico v. United States), 2004 I.C.J. 128 (Mar. 2004) (finding the United States in violation of its obligations under the Vienna Convention on Consular Relations).

106. *Simmons*, 543 U.S. at 563; *Atkins*, 536 U.S. at 321.

107. *See Atkins*, 536 U.S. at 307.

108. Penry v. Lynaugh, 492 U.S. 302, 340 (1989).

109. *Atkins*, 536 U.S. at 321.

110. *See id.* at 315-18.

111. *Id.* at 314-15. The Court noted that, in the intervening years, "the American public, legislators, scholars, and judges" had deliberated over the question and "[t]he consensus reflected in those deliberations" influenced the Court's conclusion. *Id.* at 307.

112. *Id.* at 315.

113. *Id.* at 316-17 n.21.

114. *Id.* (citing the Brief for European Union as *Amicus Curiae*). The Court found domestic support for its decision in a broad "social and professional consensus." *Id.* The Court also found international support from the overwhelming disapproval of the "world community" for the imposition of the death penalty for crimes committed by the developmentally disabled. *Id.*

115. *Id.*

religious organizations, and opinion polls."[116] Rehnquist rejected any comparative aspect to the evaluation of the "evolving standards of decency": "I fail to see, however, how the views of other countries regarding the punishment of their citizens provide any support for the Court's ultimate determination.... For if it is evidence of a *national* consensus for which we are looking, then the viewpoints of other countries simply are not relevant."[117] Justice Scalia, in his dissent (joined by Chief Justice Rehnquist and Justice Thomas), escalated the rhetorical approach of the dissent in rejecting the views of the international community: "[T]he Prize for the Court's Most Feeble Effort to fabricate 'national consensus' must go to its appeal (deservedly relegated to a footnote) to the views of assorted professional and religious organizations, members of the so-called 'world community,' and respondents to opinion polls."[118] Both dissenting opinions sought to limit or qualify the sources the Court may consider in evaluating the domestic internal discourse regarding the death penalty.[119] Moreover, by focusing exclusively on a "national consensus" and dismissing outside views "imposed" on Americans, the dissenting Justices forcefully rejected the majority's consideration of cross-cultural dialogue.[120]

In *Simmons*, the Court followed a path of analysis similar to the analysis in *Atkins* in addressing the constitutionality of the death penalty for a juvenile under age eighteen at the time the crime was committed.[121] Prior to the Court's decision in *Penry*, which allowed the execution of the developmentally disabled, in *Stanford v. Kentucky*,[122] the Court concluded that the execution of juveniles over age fifteen, but under age eighteen, was not cruel and unusual punishment.[123] In *Simmons*, the Court overruled *Stanford*.[124] As in *Atkins*, the

---

116. *Id.* at 322 (Rehnquist, C.J., dissenting).

117. *Id.* at 324-25. Rehnquist noted that earlier decisions in this area looked to international opinions, but contended that the Court had since rejected this practice. *Id.* at 325.

118. *Id.* at 347 (Scalia, J., dissenting). Scalia continued:

> Equally irrelevant are the practices of the "world community," whose notions of justice are (thankfully) not always those of our people. "We must never forget that it is a Constitution for the United States of America that we are expounding.... [W]here there is not first a settled consensus among our own people, the views of other nations, however enlightened the Justices of this Court may think them to be, cannot be imposed upon Americans through the Constitution."

Id. at 347-48 (quoting Thompson v. Oklahoma, 487 U.S. 815, 868-869 n.4 (1988) (Scalia, J., dissenting)).

119. *See id.* at 322-24 (Rehnquist, C.J., dissenting); *id.* at 347-48 (Scalia, J., dissenting).

120. *Id.* at 324-25 (Rehnquist, C.J., dissenting); *id.* at 347-48 (Scalia, J., dissenting).

121. *See* Roper v. Simmons, 543 U.S. 551, 555-56 (2005) (addressing whether it is constitutionally permissible "to execute a juvenile offender who was older than 15 but younger than 18 when he committed a capital crime").

122. *Stanford v. Kentucky*, 492 U.S. 361 (1989).

123. *See id.* at 380; *see also Thompson v. Oklahoma*, 487 U.S. 815, 830 (1988) (concluding that it would offend "standards of decency" to execute any offender under the age of sixteen at the time of the crime).

124. *Simmons*, 543 U.S. at 574.

Court began by evaluating the internal discourse regarding the execution of juveniles.[125] The Court noted that at that time, thirty states prohibited imposition of the death penalty on juveniles[126] and that the trend of legislative action since its decision in *Stanford* had consistently moved closer towards prohibition of the imposition of the death penalty on juveniles.[127] The Court in *Simmons* found this trend particularly striking in light of the "general popularity of anticrime legislation," including a trend toward "cracking down" on juvenile crime in other respects.[128] The Court in *Simmons* also considered the underlying social purposes of the death penalty, which it identified as retribution and deterrence, and concluded that neither purpose is advanced by executing juveniles, given their "diminished culpability."[129] Essentially, the Court in *Simmons* concluded that recognition of human dignity and special consideration for vulnerable groups, such as juveniles, outweigh the values of deterrence and retribution under today's standards of decency in the United States.[130]

The Court's conclusion about the standards of decency in the United States, however, is strongly buttressed through cross-cultural dialogue:

> Our determination that the death penalty is disproportionate punishment for offenders under 18 finds confirmation in the stark reality that the United States is the only country in the world that continues to give official sanction to the juvenile death penalty. This reality does not become controlling, for the task of interpreting the Eighth Amendment remains our responsibility. Yet at least from the time of the Court's decision in *Trop*, the Court has referred to the laws of other countries and to international authorities as instructive for its interpretation of the Eighth Amendment's prohibition of "cruel and unusual punishments."[131]

In the last section of the *Simmons* opinion, the Court discussed both the relevant international law and the practices of other nations.[132] The Court cited

---

125. *Id.* at 560-68.

126. This includes 12 states that reject the death penalty completely. *Id.* at 564.

127. But the trend towards prohibition of the execution of juveniles has moved at a much slower pace than the trend towards the prohibition of the execution of the developmentally disabled. *Id.* at 565 (comparing *Atkins*, the Court noted the "[i]mpressive" rate of prohibition of the death penalty for the developmentally disabled).

128. *Id.* at 566.

129. *Id.* at 571 ("Once the diminished culpability of juveniles is recognized, it is evident that the penological justifications for the death penalty apply to them with lesser force than to adults.").

130. *See id.* at 571-74. The Court concluded that "[w]hen a juvenile offender commits a heinous crime, the State can exact forfeiture of some of the most basic liberties, but the State cannot extinguish his life and his potential to attain a mature understanding of his own humanity." *Id.* at 573-74.

131. *Id.* at 575.

132. *See id.* at 575-78.

various international human rights treaties: the Convention on the Rights of the Child,[133] the International Covenant on Civil and Political Rights,[134] the American Convention on Human Rights,[135] and the African Charter on the Rights and Welfare of the Child.[136] The Court also looked to the practices of other nations and noted that since 1990, only seven other countries—Iran, Pakistan, Saudi Arabia, Yemen, Nigeria, the Democratic Republic of Congo, and China—have executed juveniles; each country has since either abolished or publicly disavowed the execution of juveniles.[137] More specifically, the Court in *Simmons* examined the United Kingdom's abolition "in light of the historic ties between our countries and in light of the Eighth Amendment's own origins," and noted that the United Kingdom abolished the use of the death penalty for juveniles over 50 years ago.[138]

Perhaps anticipating the uproar that would result from the use of international materials, the Court stated that the "opinion of the world community" does not control the Court's decision, but rather provides "respected and significant confirmation" for the Court's conclusions.[139] The Court then linked American values with international human rights standards to support its decision:

> The [Constitution] sets forth, and rests upon, innovative principles original to the American experience, such as federalism; a proven balance in political

---

133. *Id.* at 576; *see* Convention on the Rights of the Child, G.A. Res. A/44/25, at 171, U.N. GAOR, 44th Sess., U.N. Doc. A/44/49 (Nov. 20, 1989). Article 37(a) provides: "No child shall be subjected to torture or other cruel, inhuman or degrading treatment or punishment. Neither capital punishment nor life imprisonment without possibility of release shall be imposed for offenses committed by persons below eighteen years of age." *Id.* at art. 37. The United States is not a party to the Convention.

134. *Simmons*, 543 U.S. at 576; *see* ICCPR, *supra* note 10, at art. 6(5). Article 6(5) states: "Sentence of death shall not be imposed for crimes committed by persons below eighteen years of age . . . ." *Id.* The United States is a party to the Covenant, but has entered a specific reservation that purports to exclude the U.S. practice of executing juveniles:

> That the United States reserves the right, subject to its Constitutional constraints, to impose capital punishment on any person (other than a pregnant woman) duly convicted under existing or future laws permitting the imposition of capital punishment, including such punishment for crimes committed by persons below 18 years of age.

International Covenant on Civil and Political Rights, at 22, Mar. 24, 1992, S. EXEC. DOC. NO. 102-23 (1992). Interestingly, in *Simmons*, the petitioner argued that the ICCPR and the U.S. reservation supported its case. *See Simmons*, 543 U.S. at 567.

135. *Simmons*, 543 U.S. at 576; *see* ACHR, *supra* note 87. Article 4(5) provides, in part: "Capital punishment shall not be imposed upon persons who, at the time the crime was committed, were under 18 years of age or over 70 years of age . . . ." *Id.*

136. *Simmons*, 543 U.S. at 576; *see* African Charter on the Rights and Welfare of the Child art. 5(3), OAU Doc. CAB/LEG/24.9/49 (1990). Article 5(3) provides: "Death sentence shall not be pronounced for crimes committed by children." *Id.*

137. *Simmons*, 543 U.S. at 577.

138. *Id.* at 577-78.

139. *Id.* at 578.

> mechanisms through separation of powers; specific guarantees for the accused in criminal cases; and broad provisions to secure individual freedom and preserve human dignity. These doctrines and guarantees are central to the American experience and remain essential to our present-day self-definition and national identity. Not the least of the reasons we honor the Constitution, then, is because we know it to be our own. It does not lessen our fidelity to the Constitution or our pride in its origins to acknowledge that the express affirmation of certain fundamental rights by other nations and peoples simply underscores the centrality of those same rights within our own heritage of freedom.[140]

Although the dissenting opinions would have reached a different result, they offered a somewhat parallel analysis. As in *Atkins*, Justice Scalia (joined by Chief Justice Rehnquist and Justice Thomas) disagreed with the majority's assessment of the internal discourse on the death penalty for juveniles and rejected the use of cross-cultural dialogue to buttress that assessment.[141] Again Scalia argued that the "evolving standards of decency" test permits the majority to impose its own views and values upon the nation.[142] But he also raised an important point about the risk of selectivity in using international norms. Although he conflated foreign law with international human rights law, Scalia suggested that the members of the majority would reject the "reciprocity" aspect of cross-cultural dialogue in areas where it does not reflect their values.[143]

Justice O'Connor, dissenting from the majority, wrote a separate opinion to underscore her disagreement with the majority's assertion that a national consensus has emerged in opposition to the death penalty for juveniles.[144] Nevertheless, O'Connor rejected Justice Scalia's contention "that foreign and

---

140. *Id.*

141. *Id.* at 607-30 (Scalia, J., dissenting).

142. *See id.* at 607-08. Scalia stated:

> The Court thus proclaims itself sole arbiter of our Nation's moral standards—and in the course of discharging that awesome responsibility purports to take guidance from the views of foreign courts and legislatures. Because I do not believe that the meaning of our Eighth Amendment, any more than the meaning of other provisions of our Constitution, should be determined by the subjective views of five Members of this Court and like-minded foreigners, I dissent.

Id. at 608.

143. *Id.* at 624-27 (discussing foreign laws on the exclusionary rule, establishment of religion, and abortion). Scalia stated:

> The Court should either profess its willingness to reconsider all these matters in light of the views of foreigners, or else it should cease putting forth foreigners' views as part of the *reasoned basis* of its decisions. To invoke alien law when it agrees with one's own thinking, and ignore it otherwise, is not reasoned decisionmaking, but sophistry.

Id. at 627.

144. *Id.* at 604 (O'Connor, J., dissenting).

international law have no place" in considering whether the punishment is cruel and unusual under evolving standards of decency.[145] She stated:

> [T]his Nation's evolving understanding of human dignity certainly is neither wholly isolated from, nor inherently at odds with, the values prevailing in other countries. On the contrary, we should not be surprised to find congruence between domestic and international values, especially where the international community has reached clear agreement—expressed in international law or in the domestic laws of individual countries—that a particular form of punishment is inconsistent with fundamental human rights. At least, the existence of an international consensus of this nature can serve to confirm the reasonableness of a consonant and genuine American consensus.[146]

Thus, the simple framework for advancing the cultural legitimacy of human rights articulated by An-Na'im unfolded in the Supreme Court as follows: the Court evaluated domestic practices in light of the practices of other nations and examined international human rights standards to inform the understanding of domestic constitutional law and standards, which resulted in rejection of the juvenile death penalty.

B. *"Liberty presumes an autonomy of self that includes freedom of thought, belief, expression, and certain intimate conduct"*[147]*: Lessons from Lawrence about the Human Rights to Privacy and Freedom from Discrimination.*

From an international human rights law perspective, the norms regarding homosexual conduct are not as clear as those addressing the death penalty. The issue can be considered as a matter of nondiscrimination on the basis of sexual orientation or as a matter of individual or family privacy. If the issue is framed as a matter of nondiscrimination, it implicates a fundamental principle of international human rights law.[148] Nondiscrimination is highlighted in expansive terms in the International Bill of Rights instruments: the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights and the International Covenant on Economic, Social and Cultural Rights.[149] Obligations of nondiscrimination are also included in the major

---

145. *Id.* ("Over the course of nearly half a century, the Court has consistently referred to foreign and international law as relevant to its assessment of evolving standards of decency." (citations omitted)).

146. *Id.* at 605.

147. Lawrence v. Texas, 539 U.S. 558, 562 (2003).

148. The U.N. Charter includes nondiscrimination as fundamental principle of the United Nations. U.N. Charter art. 1, para. 3 (noting that one purpose of the United Nations is to promote and encourage "respect for human rights and for fundamental freedoms for all without distinction as to race, sex, language, or religion").

149. The Universal Declaration of Human Rights states: "All human beings are born free

regional human rights treaties.[150] Furthermore, two particular forms of discrimination—racial discrimination and discrimination against women—are the subjects of separate international conventions.[151] So far, there has been only one specific treaty reference to nondiscrimination on the basis of sexual orientation;[152] but much of the general treaty language prohibits discrimination on undefined "other" grounds.[153] Though the issue of discrimination based on sexual orientation has received increasing international attention, the issue remains a relatively new subject of discussion at the international level.

---

and equal in dignity and rights." UDHR, *supra* note 30, at art. 1. Article 2 provides: "Everyone is entitled to the all the rights and freedoms set forth in this Declaration, without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status." *Id.* at art. 2. The major human rights Covenants include identical language. The ICESCR provides: "The States Parties to the present Covenant undertake to guarantee that the rights enunciated in the present Covenant will be exercised without discrimination of any kind as to race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status." ICESCR, *supra* note 30, at art. 2. The ICCPR states, "Each State Party to the present Covenant undertakes to respect and to ensure to all individuals within its territory . . . the rights recognized in the present Covenant, without distinction of any kind . . . ." ICCPR, *supra* note 10, at art. 2; *see also* U.N. Charter art. 1 (stating the purpose of "promoting and encouraging respect for human rights and for fundamental freedoms for all without distinction as to race, sex, language, or religion").

150. *See* African Charter, *supra* note 24, at art. 2 (prohibiting distinction based on "race, ethnic group, colour, sex, language, religion, political or any other opinion, national and social origin, fortune, birth or other status"); ACHR, *supra* note 86, at art. 1 (prohibiting discrimination based on "race, color, sex, language, religion, political or other opinion, national or social origin, economic status, birth, or any other social condition"); European Convention, *supra* note 87, at art. 14 (prohibiting discrimination on "any ground such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status"); *see also* The Cairo Declaration on Human Rights in Islam, G.A. Res. 49/19-P, Annex, at art. 1(a), U.N. Doc A/CONF.157/PC/62/Add.18 (June 9, 1993) ("All men are equal in terms of basic human dignity and basic obligations and responsibilities, without any discrimination on the grounds of race, colour, language, sex . . . or other considerations.").

151. *See* Declaration on the Elimination of All Forms of Intolerance and of Discrimination Based on Religion or Belief, G.A. Res. 36/55, U.N. GAOR, 36th Sess., Supp. No. 51, at 171, U.N. Doc. A/36/51 (Nov. 25, 1981); CEDAW, *supra* note 56 (not ratified by the United States); International Convention on the Elimination of All Forms of Racial Discrimination, *supra* note 56 (ratified by the United States).

152. The Charter of Fundamental Rights of the European Union states: "Any discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation shall be prohibited." Charter of the Fundamental Rights of the European Union art. 21, para. 1, 2000 O.J. (C 364) 1, 13.

153. *See supra* notes 149-50 (prohibiting discrimination based on "other status" or "other social condition").

Alternatively, if the issue of sexual orientation is framed as a matter of privacy rights, the situation is similar. There are numerous human rights provisions addressing issues of personal privacy, particularly in the context of intimate and familial relationships as well as other personal liberty interests.[154] But there is no explicit language pertaining to same-sex relationships. The nations of Western Europe, whose values and history the United States views as most similar to its own, offer more explicit protection to same-sex relationships, either domestically or through the regional human rights mechanisms.[155] But the privacy and liberty protection given to same-sex relationships remains a controversy in many places around the world.[156]

Domestically, of course, the internal discourse on same-sex relationships and intimate conduct continues in the United States.[157] Interestingly, in *Lawrence v. Texas*,[158] the Court appeared to follow the internal discourse –

---

154. *See* ICCPR, *supra* note 10, at art. 23 ("The right of men and women of marriageable age to marry and to found a family shall be recognized."); ICESCR, *supra* note 30, at art. 10 ("The widest possible protection and assistance should be accorded to the family, which is the natural and fundamental group unit of society . . . ."); European Convention, *supra* note 87, at art. 8 ("Everyone has the right to respect for his private and family life, his home and his correspondence. There shall be no interference by a public authority with the exercise of this right except as in accordance with the law . . . ."); UDHR, *supra* note 30, at art. 16 ("[M]en and women of full age . . . have the right to marry and to found a family. . . . Marriage shall be entered into only with the free and full consent of the intending spouses.").

155. *See supra* note 85 (listing protections provided by Denmark, Norway, Sweden, Iceland, Finland, the Netherlands, Belgium, Spain, Germany, France, Luxembourg, Britain, Canada, Argentina, New Zealand, South Africa, Croatia, Hungary, Israel, Portugal, Slovenia, and Switzerland); *see also* Salgueiro da Silva Mouta v. Portugal, 1999-IX Eur. Ct. H.R. 309 (finding that the decision of a Portuguese court denying custody to a parent on the basis of sexual orientation discriminated against the applicant in violation of articles 8 and 14 of the European Convention).

156. "[T]he concept of family in the minds of the drafters of the Covenant was the traditional one of a man, a woman and, possibly, children. By introducing the concept of same-sex couples, Denmark was eroding the original concept and violating the basic principal of article 10." Committee on Economic, Social and Cultural Rights, *Summary Record of the 12th Meeting*, U.N. ESCOR, 20th Sess., 12th mtg. at 7, U.N. Doc. E/C.12/1999/SR.12 (May 6, 1999) (comments by Mr. Sadi). Canada's Parliament established a committee to research the issue; the committee determined "that the definition of marriage as the voluntary union between one man and one woman to the exclusion of all others was no longer acceptable under Canadian law." Committee on the Rights of the Child, *Summary Record of the 895th Meeting*, ¶ 61, U.N. Doc. CRC/C/SR.895 (Sept. 17, 2003) (statement of Mr. Farber of Canada). Same-sex couples were already raising children and, therefore, "the legalization of same-sex unions would provide such children with the same rights as those enjoyed by children from traditional marriages." *Id.*

157. The discourse regarding privacy interests and the prohibition of discrimination against same-sex relationships is at an earlier stage than the discourse concerning the death penalty. Religion and culture play powerful roles as the debate takes shape. But even in this context, human rights advocates are using values widely accepted in the United States, including nondiscrimination and privacy, to argue for a human rights-oriented approach.

158. 539 U.S. 558 (2003).

cross-cultural dialogue construct in its rationale more closely than it did in the death penalty cases. As in *Atkins* and *Simmons*, the Court evaluated the history of the debate, noted changing public views, and overruled a previous decision to reach a human rights-based result.[159] Once again, the Court created controversy by looking outside the United States to examine the practices of other nations and the perspectives of the international community.[160]

The opinion in *Lawrence* began by re-characterizing the rights at stake.[161] In the 1986 case, *Bowers v. Hardwick*,[162] the Court upheld a Georgia statute criminalizing certain same-sex intimate conduct, finding that the Constitution does not confer "a fundamental right upon homosexuals to engage in sodomy."[163] Almost twenty years later in *Lawrence*, the Court invalidated a similar Texas statute, framing the issue as a matter of private conduct in the "exercise of . . . liberty" under the Fourteenth Amendment's Due Process Clause.[164] After re-characterizing the rights at stake in this more generous manner, the Court undertook a reinterpretation of the historical treatment of sodomy it relied upon in *Bowers*.

The Court in *Bowers* had quashed any sense of internal discourse on the issue by connecting the criminalization of sodomy to "ancient roots" and presenting a picture of a long and unchanged historical condemnation of consensual sodomy.[165] In *Lawrence*, the Court destabilized that notion by carefully reexamining its history and reinterpreting sodomy prohibitions as general condemnations of non-procreative sex, rather than as established traditions of prosecuting homosexual conduct.[166] The general condemnation of

---

159. *Id.* at 578-79.

160. *Id.* at 572-73. "To the extent *Bowers* relied on values we share with a wider civilization, it should be noted that the reasoning and holding in *Bowers* have been rejected elsewhere." *Id.* at 576.

161. *Id.* at 562 (noting that the case "involves liberty of the person both in its spatial and in its more transcendent dimensions").

162. 478 U.S. 186 (1986).

163. *Id.* at 190, 196.

164. *Lawrence*, 539 U.S. at 564. The Court characterized the liberty interest broadly as "liberty of the person both in its spatial and in its more transcendent dimensions." *Id.* at 562. The Court explained:

> The laws involved in *Bowers* and here are, to be sure, statutes that purport to do no more than prohibit a particular sexual act. Their penalties and purposes, though, have more far-reaching consequences, touching upon the most private human conduct, sexual behavior, and in the most private of places, the home. The statutes do seek to control a personal relationship that, whether or not entitled to formal recognition in the law, is within the liberty of persons to choose without being punished as criminals.

Id. at 567.

165. *Bowers*, 478 U.S. at 192-94.

166. *Lawrence*, 539 U.S. at 570-74. The Court stated that it was only in the 1970s that states began singling out same-sex relations for criminal prosecution. *Id.* at 570 (noting that only nine states have criminalized same-sex relations). The Court also noted that some of those states have since moved towards abolishing the prohibitions. *Id.*

non-procreative sex has been the topic of much national debate; moreover, the private, intimate conduct of individuals has been subject to increasing protection by the Court for the past fifty years.[167]

In *Lawrence*, the Court openly addressed the importance of the social values at stake in this debate, particularly in the context of same-sex intimate conduct:

> The condemnation has been shaped by religious beliefs, conceptions of right and acceptable behavior, and respect for the traditional family. For many persons these are not trivial concerns but profound and deep convictions accepted as ethical and moral principles to which they aspire and which thus determine the course of their lives.[168]

But the Court determined that recent laws and traditions are most relevant to defining the accepted scope of liberty.[169] The Court found "an emerging awareness that liberty gives substantial protection to adult persons in deciding how to conduct their private lives in matters pertaining to sex."[170] After reviewing the current laws regarding intimate conduct, as in the death penalty cases, the Court relied upon the trend by individual states either to repeal laws criminalizing same-sex intimate conduct or to decline enforcement of those laws.[171] The Court emphasized the powerful impact of such laws and their resulting prosecutions on human dignity and the fundamental importance of the liberty interests at stake.[172]

---

167. The Court traced this history, beginning with *Griswold v. Connecticut*, 381 U.S. 479 (1965) (invalidating a state law prohibiting the sale or use of contraception on grounds that the law infringed upon the right to privacy in marriage), *Eisenstadt v. Baird*, 405 U.S. 438 (1972) (extending *Griswold* to unmarried persons), *Roe v. Wade*, 410 U.S. 113 (1973) (holding that the right to an abortion is protected by liberty interests in Due Process Clause), and *Carey v. Population Services Int'l*, 431 U.S. 678 (1977) (invalidating a law prohibiting sale of contraceptives to persons under 16 years of age). The Court also cited to more recent decisions, including *Planned Parenthood of Southeastern Pa. v. Casey*, 505 U.S. 833 (1992) (reaffirming the constitutional protection of personal decisions regarding marriage, procreation, contraception, and family relationships), and *Romer v. Evans*, 517 U.S. 620 (1996) (striking down class-based legislation directed at homosexuals).

168. *Lawrence*, 539 U.S. at 571.

169. *Id.* at 571-72.

170. *Id.* at 572.

171. *Id.* at 573. The Court noted that the number of states with anti-sodomy laws dropped from twenty-five at the time of its decision in *Bowers* to thirteen at the time of its decision in *Lawrence*. In addition, only four of those states enforced their laws exclusively against homosexual intimate conduct. *Id.* Even in states with sodomy laws, there was a pattern of non-enforcement in cases involving consenting adults acting in private. *Id.*

172. *Id.* at 575-76. The court expressed concern about the consequences of such laws, including public and private discrimination, stigma, risk of prosecution, and collateral consequences of conviction. *Id.*

Nevertheless, the Court in *Lawrence* did not limit itself to consideration of the evolving national discourse. In its reevaluation of the historical and social context, the Court also included cross-cultural perspectives.[173] It explained that the British Parliament had considered repealing laws punishing homosexual conduct as early as 1957 and took steps in that direction in the 1960s.[174] The Court also mentioned the European Court of Human Rights' 1981 decision in *Dudgeon v. United Kingdom*[175] and subsequent case law following its precedent.[176] In *Dudgeon*, the European Court found that the laws of Northern Ireland prohibiting homosexual conduct violated the European Convention on Human Rights.[177] Noting that this "integral part of human freedom" has been recognized in many other countries, the Supreme Court concluded that it could not find justification for giving it less protection in the United States.[178]

The majority decision in *Lawrence* used the language of human rights and human dignity, reframed the historical perspective to more fully reflect the ongoing internal discourse, and cited to cross-cultural developments protecting the interests at issue.[179] In this way, the Court has recognized and advanced the cultural legitimacy of human rights norms protecting privacy and prohibiting discrimination for same-sex relationships. Moreover, despite their different outcomes, both the concurring and dissenting opinions also fit within this methodological framework.

The concurrence by Justice O'Connor framed the issue in the case as a matter of equal protection, rather than liberty and due process.[180] O'Connor used the familiar language of the equal protection analysis to make an argument that reflects fundamental principles of nondiscrimination: "I am confident,

---

173. The Court explained, "[w]hen our precedent has been thus weakened, criticism from other sources is of greater significance." *Id.* at 576.

174. *Id.* at 572-73.

175. 45 Eur. Ct. H.R. (1981).

176. The Supreme Court noted that the European Court of Human Rights continued to follow its precedent in *Dudgeon* rather than giving consideration to the Supreme Court's reasoning in *Bowers*. *Lawrence*, 539 U.S. at 576.

177. *Dudgeon*, 45 Eur. Ct. H.R. ¶ 52. The Supreme Court in *Lawrence* pointed out that this decision is authoritative for all forty-five nations that are members of the Council of Europe. *Lawrence*, 539 U.S. at 573.

178. *Id.* at 577.

179. The *Lawrence* opinion referenced numerous amici curiae, representing a wide range of perspectives. Amici included the Cato Institute, the American Civil Liberties Union, and Mary Robinson (former U.N. High Commissioner for Human Rights). *Id.* at 568, 576-77.

180. Likewise, the majority briefly addressed the issue as a matter of equal protection. It noted the connection between equal treatment and the due process right to "demand respect for conduct protected by the substantive guarantee of liberty." *Id.* at 575. But the Court was concerned that invalidation based on equal protection grounds would not go far enough and would be vulnerable to more carefully crafted laws. *Id.* The majority also determined that equal protection arguments would not address the important issues of stigma or dignity. *Id.* Further, it stated that the continued existence of *Bowers* as precedent "demeans the lives of homosexual persons." *Id.*

however, that so long as the Equal Protection Clause requires a sodomy law to apply equally to the private consensual conduct of homosexuals and heterosexuals alike, such a law would not long stand in our democratic society."[181] But O'Connor also noted that social values, such as national security and preserving the traditional institution of marriage, might be sufficient to overcome the value of equal treatment in other circumstances.[182]

In his dissent, Justice Scalia emphasized a contrasting view of the internal national discourse and a fierce reaction to cross-cultural dialogue using stronger language than he used in the death penalty cases. He framed the issue squarely as a "culture war," and he used this characterization to reject the Court's involvement in advancing or determining that cultural evolution:

> It is clear from this that the Court has taken sides in the culture war, departing from its role of assuring, as neutral observer, that the democratic rules of engagement are observed. Many Americans do not want persons who openly engage in homosexual conduct as partners in their business, as scoutmasters for their children, as teachers in their children's schools, or as boarders in their home. They view this as protecting themselves and their families from a lifestyle that they believe to be immoral and destructive. The Court views it as "discrimination" which it is the function of our judgments to deter. So imbued is the Court with the law profession's anti-anti-homosexual culture, that it is seemingly unaware that the attitudes of that culture are not obviously "mainstream". . . .[183]

Scalia also strongly rejected the cross-cultural perspective incorporated by the majority opinion as "meaningless" but "[d]angerous dicta": "Constitutional entitlements do not spring into existence because some States choose to lessen or eliminate criminal sanctions on certain behavior. Much less do they spring into existence, as the Court seems to believe, because *foreign nations* decriminalize conduct."[184]

In *Lawrence*, the Court followed the conceptual internal discourse – cross-cultural dialogue framework, but their analysis unfolded somewhat differently than in the death penalty cases. The Court focused primarily on reinterpreting the domestic history and discourse on intimate conduct, as broadly defined.[185] The court used the language of human rights and human dignity, and it drew upon both the international human rights law developed by the European Court

---

181. *Id.* at 584-85 (O'Connor, J., concurring).

182. *Id.* at 585.

183. *Id.* at 602-03 (Scalia, J., dissenting). Scalia later clarified:

> Let me be clear that I have nothing against homosexuals, or any other group, promoting their agenda through normal democratic means. Social perceptions of sexual and other morality change over time, and every group has the right to persuade its fellow citizens that its view of such matters is the best.

Id. at 603.

184. *Id.* at 598 (emphasis added).

185. *Id.* at 562-571 (majority opinion).

of Human Rights and the domestic practice of the United Kingdom.[186] In this way, the Court both recognized and advanced the cultural legitimacy of human rights norms protecting privacy and prohibiting discrimination. But because the international human rights norms at stake are less explicit and more contested both domestically and globally, the value of the cross-cultural perspectives is more problematic. As a result, the cultural legitimacy of the decision—and the human rights values it advances—is perhaps more tenuous.

C. Whose Law Will It Be? Looking Forward with the New Court.

Although the recent decisions in *Atkins*, *Lawrence,* and *Simmons* were authored by sitting members of the Court,[187] the Supreme Court has undergone a recent change in composition. With the retirement of Justice O'Connor and the death of Chief Justice Rehnquist, the membership of the Court now includes new Chief Justice Roberts and Justice Alito.[188] The internal discourse on the propriety of the Court's consideration of international human rights standards was evident in both men's confirmation hearings. Former Judge Roberts responded to questioning on the issue from Senator Coburn:

> [Senator COBURN.] My question to you is, relying on foreign precedent and selecting and choosing a foreign precedent to create a bias outside of the laws of this country, is that good behavior?
> Judge ROBERTS. Well, I . . . don't think it's a good approach. I wouldn't accuse judges or Justices who disagree with that, though, of violating their oath. I'd accuse them of getting it wrong on that point, and I'd hope to sit down with them and debate it and reason about it.
> . . . .
> Senator COBURN. Can the American people count on you to not use foreign precedent in your decision making on the Supreme Court?
> Judge ROBERTS. You know, I will follow the Supreme Court's precedents consistent with the principles of *stare decisis*, and there are cases in this area,

---

186. *Id.* at 572-73.

187. Justice Stevens authored the 2002 opinion in *Atkins v. Virginia*, joined by Justices O'Connor, Kennedy, Souter, Ginsburg, and Breyer. Atkins v. Virginia, 536 U.S. 304, 305 (2002). Chief Justice Rehnquist and Justices Scalia and Thomas dissented. *Id.* Justice Kennedy authored both the 2003 opinion in *Lawrence v. Texas* and the 2005 decision in *Roper v. Simmons* and was joined by Justices Stevens, Souter, Ginsburg, and Breyer in both decisions. Roper v. Simmons, 543 U.S. 551, 554 (2005); *Lawrence*, 539 U.S. at 561. Justice O'Connor concurred in the result in *Lawrence* and dissented in *Simmons*; Chief Justice Rehnquist and Justices Scalia and Thomas dissented in both cases. *Simmons*, 543 U.S. at 554; *Lawrence*, 539 U.S. at 561.

188. The Senate confirmed John Roberts by a vote of 78-22, and he was sworn in as Chief Justice of the Supreme Court on September 29, 2005. Charlie Savage, *Roberts Becomes Nation's 17th Chief Justice*, BOSTON GLOBE, Sept. 30, 2005, at A1, *available at* 2005 WLNR 15425224. After being confirmed by the Senate by a vote of 58-42, Samuel Alito Jr. was sworn in on January 31, 2006. Glenn Thrush, *Amid Bitter Split, Senate Oks Alito*, NEWSDAY, Feb. 1, 2006, at A02, *available at* 2006 WLNR 1775822.

> of course. That's why we're having the debate. The Court has looked at those. I think it's fair to say, in the prior opinions, those are not determinative in the sense that the precedent turned entirely on foreign law, so it's not a question of whether or not you'd be departing from these cases if you decided not to use foreign law.[189]

Roberts also had a lengthy exchange with Senator Kyl on the issue.[190] Senator Kyl specifically referred to *Roper v. Simmons* in his questioning.[191] A few months later, then Judge Alito faced similar questions and he was even more explicit in his response:

> [Senator KYL.] What is the proper role, in your view, of foreign law in U.S. Supreme Court decisions, and when, if ever, is citation to or reliance on these foreign laws appropriate?
> Judge ALITO. I don't think that foreign law is helpful in interpreting the Constitution. Our Constitution does two basic things. It sets out the structure of our Government and it protects fundamental rights. The structure of our Government is unique to our country, and so I don't think that looking to decisions of supreme courts of other countries or constitutional courts in other countries is very helpful in deciding questions relating to the structure of our government.
>
> As for the protection of individual rights, I think that we should look to our own Constitution and our own precedents. . . .
>
> We have our own law, we have our own traditions, we have our own precedents, and we should look to that in interpreting our Constitution. . . .
>
> . . . .
>
> [Senator COBURN.] The question I have for you—and I could not get Judge Roberts to answer it because of the conflict that might occur afterwards, but I have the feeling that the vast majority of Americans do not think it is proper for the Supreme Court to use foreign law. . . . I just wondered if you had any comments on that comment.
> Judge ALITO. Well, I don't think that we should look to foreign law to interpret our own Constitution. . . . I think the Framers would be stunned by the idea that the Bill of Rights is to be interpreted by taking a poll of the countries of the world. . . . The Framers did not want Americans to have the rights of people in France or the rights of people in Russia, or any of the other countries on the continent of Europe at the time. They wanted them to have the rights of Americans, and I think we should interpret our Constitution—we should interpret our Constitution. I don't think it's appropriate to look to foreign law.[192]

---

189. *Roberts Confirmation Hearing*, *supra* note 18.

190. *Id.*

191. Although Judge Roberts declined to comment on the specific case, he raised general concerns about the use of "foreign law" as contrary to democratic theory and as improperly expanding judicial discretion. *Id.*

192. *Alito Confirmation Hearing*, *supra* note 3. Alito also commented on the issue in response to questioning from Senators Leahy (affirming that English common law may help in

Both nominees were firm in their renunciation of the use of "foreign" law sources. This approach would accord with that of former Chief Justice Rehnquist, but differ from that of former Justice O'Connor.[193] The other members of the Court are split in their attitudes toward the use of international law: Justices Kennedy, Stevens, Souter, Ginsburg and Breyer[194] have referred to international law and will likely continue to do so; Justices Scalia and Thomas[195] have been adamantly opposed to such use and will likely continue to be. Thus, one would expect that the addition of Chief Justice Roberts and Justice Alito will slightly shift the Court's willingness to consider, and certainly to cite, international and foreign law sources.

But it is not just the addition of new members to the Court that may change its willingness to consider international law and the practices of other nations. The public interest, and often public outcry, on this issue is likely to have an effect.[196] The prominent questioning of the nominees by members of Congress,

---

understanding United States law) and Kohl (clarifying that he does not think it is proper to look to foreign law in interpreting the constitution, but that it may be helpful in looking to the practices of foreign countries in how they organize their constitutional courts). *Id.*

193. Chief Justice Rehnquist was explicit in his views in *Atkins v. Virginia*:

> I write separately, however, to call attention to the defects in the Court's decision to place weight on foreign laws, the views of professional and religious organizations, and opinion polls in reaching its conclusion. The Court's suggestion that these sources are relevant to the constitutional question finds little support in our precedents and, in my view, is antithetical to considerations of federalism. . . .

536 U.S. 304, 322 (2002) (Rehnquist, C.J., dissenting) (citation omitted). In contrast, Justice O'Connor wrote separately in *Roper v. Simmons* to affirm her view that it was appropriate for the Court to consider foreign and international law: "I disagree with Justice Scalia's contention that foreign and international law have no place in our Eighth Amendment jurisprudence. Over the course of nearly half a century, the Court has consistently referred to foreign and international law as relevant to its assessment of evolving standards of decency." 543 U.S. 551, 604 (2005) (O'Connor, J., dissenting) (citation omitted).

194. *See Simmons*, 543 U.S. at 554; *Lawrence*, 539 U.S. at 561; *Atkins*, 536 U.S. at 305; *see also* Jeffrey Toobin, *Swing Shift: How Anthony Kennedy's Passion for Foreign Law Could Change the Supreme Court*, THE NEW YORKER, Sept. 12, 2005, at 42 (describing Justice Kennedy's "passion" for foreign and international law); Justice Ruth Bader Ginsburg, Remarks at the Annual Meeting of the American Society of International Law, "A Decent Respect to the Opinions of [Human]kind": The Value of a Comparative Perspective in Constitutional Adjudication (Apr. 1, 2005), *available at* http://www.asil.org/events/AM05/ginsburg050401.html ("If U.S. experience and decisions can be instructive to systems that have more recently instituted or invigorated judicial review for constitutionality, so we can learn from others now engaged in measuring ordinary laws and executive actions against charters securing basic rights.").

195. *See Simmons*, 543 U.S. at 607 (Scalia, J., and Thomas, J., dissenting); *Lawrence*, 539 U.S. at 586 (Scalia, J., and Thomas, J., dissenting); *Atkins*, 536 U.S. at 321 (Scalia, J., and Thomas, J., dissenting); *see also* Lane, *supra* note 4, at A1 (accounting the debate between Justice Antonin Scalia and Justice Stephen G. Breyer at American University Washington College of Law over the Supreme Court's use of foreign law in its own decisions).

196. In fact, the use of these cross-cultural sources has become a part of the broader

the proposed Congressional Resolution to prohibit the use of foreign sources by the Court, and the public statements by members of the administration[197] all suggest that the easiest course for the Court would be to refrain from further invocation of international law, practice, or norms. That would seem to be the politically wise choice, and given the supplementary part these sources have played in the decisions themselves, a choice with little real cost. Yet, such a conclusion ultimately reflects a misunderstanding of the importance of the dynamic at work in this process. The Court has an important function in both the internal, domestic discourse on human rights norms and in the ongoing cross-cultural dialogue on those norms. It remains to be seen what roles Chief Justice Roberts and Justice Alito will assume in that process, but it is unlikely that the Court will abandon its responsibility to participate in and stimulate dialogue completely.

CONCLUSION

The bounds and content of the relationship between the United States and other nations in the international community always provide interesting, and often provocative, questions. Equally provocative, at both the domestic and international levels, is the role that "outsiders" should play in changing the nature of a domestic legal or cultural system. The United States appears increasingly willing to be that outsider by providing guidance or insisting on the implementation of international legal obligations, including human rights obligations, in the domestic systems of other nations. The United States has been less willing to re-examine its own domestic laws and policies, viewing cross-cultural or international input as that of outsiders who are imposing their views on the nation; many in the nation believe that accepting outsider input would offend national sovereignty and be inconsistent with domestic views and values.

Does this resistance reflect the success of efforts to change international human rights standards to be more reflective of other cultural perspectives? If the controversial cases revolved around economic, social, and cultural rights, the assertion may be persuasive. But the decisions on the death penalty and sodomy laws involve basic civil and political rights, which have historically been viewed as most compatible with U.S. notions of individual rights. Is the rejection of international sources a reflection of a post-9/11 United States that views any outside involvement as suspect? It is true that there may be a more isolationist sentiment in the United States now, but the current debate about the death penalty long preceded the events of 9/11, and the controversy around same-sex relationships has no apparent connection to concerns about terrorism and national security.

---

internal discourse on the substance of the human rights norms at issue in the cases. *See supra* Part I.B.

197. *See supra* notes 4, 28 and accompanying text.

Perhaps Professor An-Na'im's suggestions, which focus on the importance of the domestic cultural (and religious) context at any given time, provide a simple approach for examining a complex problem. American "exceptionalism" on this issue is not exceptional at all. The recent Supreme Court decisions—*Atkins*, *Simmons*, and *Lawrence*—suggest that there may be a growing openness to an internal re-examination on human rights protections in the United States and to a cross-cultural dialogue on these issues. But the outcry in response to the proposed use of human rights law suggests reasons to remain cautious. The confirmation hearings of both new Justices have included direct questioning on these decisions and the role of foreign and international law in Supreme Court decisions.

From a normative and legal human rights perspective, the significant issue is the development of cultural legitimacy, rather than purely formal implementation. If limiting or abolishing the death penalty is viewed as integral to a system of American justice and fundamental for the protection of the right to life, then it will be accepted in the United States. If recognizing the privacy of intimate relationships between persons of the same sex is considered a logical extension of the existing protection for intimate relationships and a commitment to equal treatment, then the political support will exist to ensure that liberty. If the values reflect the culture of the United States and can serve as both a model to and reflection of the values of the broader global community, then the laws and the implementation of those laws, both domestic and international, will follow.

PROPERTY LAW—WILLS—EFFECT OF LAPSED RESIDUARY GIFTS IN THE STATE OF TENNESSEE

In re Estate of McFarland, 167 S.W.3d 299 (2005).

I. INTRODUCTION

Ms. Merle Jeffers McFarland executed a holographic will on November 14, 1994.[1] The will reserved two percent of her estate for her funeral expenses and left three thousand dollars to the Tieke-McCullough Cemetery bank fund.[2] A residuary clause split the remainder between eighteen individuals and organizations,[3] specifying the percentage that she wished to bequest to each particular person or entity.[4] Ms. McFarland died on October 12, 2001.[5] However, three of the residuary beneficiaries predeceased her, and no issue survived any of them.[6]

The Chancery Court for Hawkins County appointed an administrator and admitted the will to probate.[7] The administrator filed a declaratory judgment action seeking direction for disbursement of the predeceased beneficiaries'

---

1. *In re* Estate of McFarland, 167 S.W.3d 299, 301 (Tenn. 2005). A holographic will is "[a] will that is entirely handwritten by the testator." BLACK'S LAW DICTIONARY 1592 (7th ed. 1999). "In many states a holographic will is valid even if not witnessed." *Id.* Holographic wills are valid in roughly half of the states, mainly in the West. John H. Langbein, *Excusing Harmless Errors in the Execution of Wills: A Report on Australia's Tranquil Revolution in Probate Law*, 87 COLUM. L. REV. 1, 3 (1987). Holographic wills are valid in Tennessee. *See* TENN. CODE ANN. § 32-1-105 (2001).

2. *McFarland*, 167 S.W.3d at 301. The will also specified an estate administrator and provided burial instructions. *Id.*

3. *Id.* The residuary clause was introduced with the following statement: "[t]he rest of the estate I wish to be divided to the following." *Id.* A residuary clause is defined as "[a] testamentary clause that disposes of any estate property remaining after the satisfaction of specific bequests and devises." BLACK'S LAW DICTIONARY 1311 (7th ed. 1999).

4. *McFarland*, 167 S.W.3d at 301. Ms. McFarland left ten percent to each of her two brothers, Willie Lee Jeffers and Minnis Rankin Jeffers; ten percent each to Clarence Lee McFarland, Mary Louise McFarland, and Evelyn B. McFarland McCulley; ten percent divided equally between Clyde E. McFarland's three sons; five percent to the First United Methodist Church of Bulls Gap; five percent to Larry and Virginia Carpenter; two percent to the city of Bulls Gap; two percent to the Tieke-McCullough Cemetery for upkeep; ten percent to the Thompson Cancer Center for research; ten percent to the University of Tennessee for scholarships and other needs; one percent each to the Bulls Gap Masonic Lodge and the Easter Star Lodge; and two percent to the United Way Fund or "any other worthy" charity. *Id.*

5. *Id.*

6. *Id.* at 302. The predeceased beneficiaries were Minnis Rankin Jeffers, Willie Lee Jeffers, and Mary Louise McFarland. *Id.*

7. *Id.* at 301.

funds.[8] The court found that the gifts had lapsed.[9] Because each of the predeceased beneficiaries had died without issue, the Tennessee anti-lapse statute was inapplicable.[10] Applying the common law rule set forth in *Ford v. Ford*[11] for determining the beneficiary of a lapsed residuary gift,[12] the trial court concluded that the testatrix had died partially intestate.[13] Therefore, the chancellor directed that the lapsed portion would not pass to the remaining residuary beneficiaries; instead, it would pass to the testatrix's heirs at law[14] as prescribed by the Tennessee intestacy statute.[15]

On interlocutory appeal,[16] the Tennessee Court of Appeals "adhere[d] to the clear precedent" of *Ford* and affirmed the trial court's decision.[17] Seeking de novo review,[18] the administrator of the estate argued that the lapsed bequests should have passed to the remaining residuary beneficiaries in proportion to their designated shares,[19] as proposed by the Uniform Probate Code's "modern" rule.[20] The appellee, the heir at law, agreed with the trial court and the court of appeals that the lapsed bequests should have passed according to the Tennessee intestacy statute.[21] On appeal to the Supreme Court of Tennessee, *held*,

---

8. *Id.* at 301. The appellant estate administrator was Stephen D. McFarland. *See* Brief of Appellant at vi, In re Estate of McFarland, 167 S.W.3d 299 (Tenn. 2005) (No. E2003-01833-SC-SO9-CV).

9. *McFarland*, 167 S.W.3d. at 302. "Lapse" is defined as "[t]he failure of a testamentary gift, esp[ecially] when the beneficiary dies before the testator dies." BLACK'S LAW DICTIONARY 885 (7th ed. 1999).

10. *See* TENN. CODE ANN. § 32-3-105(a) (2001); *see also* text accompanying note 53.

11. 31 Tenn. 431 (1852).

12. *See id.* at 433; *see also* text accompanying note 85.

13. *McFarland*, 167 S.W.3d at 302. As the appellant's brief notes, this partial intestacy amounted to thirty percent of Ms. McFarland's estate. Brief of Appellant at vi, In re Estate of McFarland, 167 S.W.3d 299 (Tenn. 2005) (No. E2003-01833-SC-SO9-CV).

14. *McFarland*, 167 S.W.3d at 302.

15. *See generally* TENN. CODE ANN. §§ 31-2-101 to -110 (2001) (detailing the applicable intestate succession provisions).

16. "[A]n appeal by permission may be taken from an interlocutory order of a trial court from which an appeal lies to the Supreme Court, Court of Appeals or Court of Criminal Appeals only upon application and in the discretion of the trial and appellate court." TENN. R. APP. P. 9(a). Despite its designation as "appellee" in the Court of Appeals opinion, the University advocated the appellant's position. *In re* Estate of McFarland, No. E2003-01833-COA-R9CV, 2004 WL 2399733, at *1 n.2 (Tenn. Ct. App. Oct. 27, 2004).

17. In re Estate of McFarland, No. E2003-01833-COA-R9CV, 2004 WL 2399733, at *3 (Tenn. Ct. App. Oct. 27, 2004). *See also McFarland*, 167 S.W.3d at 302 (discussing the appellate court's affirmation of the trial court decision).

18. Tennessee courts will review probate cases de novo when there is no factual dispute. *McFarland*, 167 S.W.3d at 302 (citing In re Estate of Vincent, 98 S.W.3d 146, 148 (Tenn. 2003)).

19. *McFarland*, 167 S.W.3d at 302.

20. *See infra* text accompanying notes 75-77.

21. *McFarland*, 167 S.W.3d at 302. Appellee, the heir at law, was James Cox. *See* Answer of Appellee at 2, In re Estate of McFarland, 167 S.W.3d 299 (Tenn. 2005) (No. E2003-

affirmed.[22] Where there is no evidence of the testator's intent, lapsed residuary gifts result in partial intestacy and pass to the testator's heirs at law, not to remaining residuary beneficiaries, because (1) will construction principles presume the testator's acquaintance with the *Ford* rule when executing a will and (2) the doctrine of stare decisis requires Tennessee courts to follow the longstanding *Ford* rule. *In re Estate of McFarland*, 167 S.W.3d 299 (2005).

II. DISTRIBUTION OF LAPSED RESIDUARY GIFTS

For over 150 years, the *Ford* holding remained unscathed.[23] The Tennessee Supreme Court never revisited the issue, and the Tennessee Court of Appeals only applied the rule once.[24] In 2005, however, the Tennessee Supreme Court granted review in *In re Estate of McFarland*.[25] By that time, forty-one states had adopted the modern rule regarding lapsed residuary gifts either by statute or judicial decision.[26] Thus, the court faced the problem of appropriately distributing a lapsed residuary gift for the first time in over 150 years.[27] The Tennessee Supreme Court had two options.[28] It could have overruled the *Ford* decision and followed other states by adopting the Uniform Probate Code's (UPC) modern rule.[29] Alternatively, the court could have upheld *Ford* and rejected the UPC's modern rule. The court chose to reject the UPC's modern rule and held that the lapsed gift resulted in partial intestacy and would pass to the testatrix's heirs at law.[30]

A. Will Construction Principles

The most basic duty of the courts in construing wills is to determine the intent of the testator and carry out the testator's wishes.[31] The Tennessee

---

01833-SC-SO9-CV).

22. *McFarland*, 167 S.W.3d at 301.

23. *See id.* at 306 (Drowota, C.J., dissenting).

24. *Id.* at 308. *Davis v. Anthony* is the only reported Tennessee Court of Appeals case that has applied the *Ford* rule. *See* Davis v. Anthony, 384 S.W.2d 60, 63 (Tenn. Ct. App. 1964).

25. *See McFarland*, 167 S.W.3d at 301.

26. *Id.* at 304 n.5.

27. *Id.* at 306 (Drowota, C.J., dissenting).

28. *Id.* at 302 (majority opinion).

29. *Id.* at 304 n.5.

30. *McFarland* 167 S.W.3d at 301.

31. Winningham v. Winningham, 966 S.W.2d 48, 50 (Tenn. 1998) (quoting In re Walker, 849 S.W.2d 766, 768 (Tenn. 1993)). *See also, e.g.*, Cowden v. Sovran Bank/Cent. S., 816 S.W.2d 741, 744 (Tenn. 1991) ("The cardinal and basis rule in the construction of wills is that the court shall seek to discover the intention of the testator, and will give effect to it unless it contravenes some rule of law or public policy." (quoting Bell v. Shannon, 367 S.W.2d 761, 766 (Tenn. 1963)).

Supreme Court emphasized this duty as early as 1836 in *Williams v. Williams*,[32] when it stated, "[I]n the construction of last wills and testaments, the scope and import of the entire instrument are to be considered, for the purpose of discovering the intention of the testator, and that such intention, when discovered, is of paramount and controlling influence."[33] Thus, the testator's intent should be gleaned "from the particular words used, from the context, and from the general scope and purpose of the instrument,"[34] not by "mere surmise."[35] Holographic wills are to be construed more liberally, however, because they are "not drafted by . . . skilled drafts[men]."[36]

On rare occasions, Tennessee courts will not adhere to the testator's intent.[37] The courts will not carry out the testator's wishes if they "contravene[] some rule of law or public policy."[38] For example, wills that violate the rule against perpetuities are invalid, regardless of the testator's intent.[39]

In addition to examining the language of the will itself, the law requires Tennessee courts to make three general presumptions when determining a testator's intent.[40] First, a court must read a will as if it were executed immediately before the testator's death.[41] This presumption is codified in Tennessee Code Annotated § 32-3-101.[42] Inspired by the English Wills Act of 1837, the Tennessee legislature adopted this statute in 1842, and it has been "on the books" ever since.[43] Second, a court will assume that the testator did not intend to die intestate—even partially intestate.[44] The Tennessee Supreme Court incorporated this presumption against intestacy into case law.[45] The opinion in *Williams v. Williams*[46] states that a testator "is not to be presumed as intending, with reference to any portion of his property, to die intestate."[47]

---

32. 18 Tenn. 20 (1836).

33. *Id.* at 25.

34. Hoggatt v. Clopton, 217 S.W. 657, 659 (Tenn. 1919) (quoting Lynch v. Burts, 48 Tenn. 600, 604 (1870)).

35. Pinkerton v. Turman, 268 S.W.2d 347, 350 (Tenn. 1954).

36. Garner v. Becton, 212 S.W.2d 890, 891 (Tenn. 1948).

37. *See* Bell v. Shannon, 367 S.W.2d 761, 765-66 (Tenn. 1963).

38. *d.* at 766. As a general matter, courts will not give effect to the testator's intent if "the results would be absurd or against public policy as expressed in state and federal Constitutions, statutes, and judicial decisions." 80 AM. JUR. 2D WILLS § 1010 (2002).

39. *See generally, e.g.*, Sands v. Fly, 292 S.W.2d 706 (Tenn. 1956) (recognizing that devises violating the rule against perpetuities are void in Tennessee).

40. *See* TENN. CODE ANN. § 32-3-101 (2001).

41. *Id.*

42. *See id.*

43. *ee infra* note 61.

44. *Id.*

45. *See* Williams v. Williams, 18 Tenn. 20, 20 (1836).

46. 8 Tenn. 20 (1836).

47. *Id.* at 20.

Courts must also presume that the testator was familiar with the law governing probate.[48] This presumption of knowledge of the law is not unique to matters of will interpretation,[49] as it applies to all areas of Tennessee law.[50] In *McCarley v. McCarley*,[51] however, the Tennessee Supreme Court explicitly applied this rule to the execution of wills.[52] The court presumed that the testatrix was aware that the law considered her sons capable of begetting children until they died, regardless of how long they lived.[53]

B. The Development of Tennessee's "Traditional" Rule and the Uniform Probate Code's "Modern" Rule

Sometimes it is nearly impossible to implement the intent of the testator. Generally, when a will bequests a gift to someone who predeceases the testator, the gift lapses.[54] To prevent this outcome, the Tennessee General Assembly enacted an antilapse statute in 1842.[55] Under this statute, a deceased beneficiary's issue receives the otherwise lapsed gifts.[56] Without the statute, the testator would die partially intestate, and the gift would pass to the testator's heirs at law, who may or may not be named in the will.[57] Presumably, this is an undesirable result. Accordingly, the statute seeks to carry out the intent of the testator.[58]

The current version of Tennessee's antilapse statute is found at Section 32-3-105[59] of the Tennessee Code Annotated. Sub-section (a) reads:

> Whenever the devisee or legatee or any member of a class to which an immediate devise or bequest is made, dies before the testator, or is dead at the making of the will, leaving issue which survives the testator, the issue

---

48. McCarley v. McCarley, 360 S.W.2d 27, 29 (Tenn. 1962) (quoting Latta v. Brown 34 S.W. 417, 419 (Tenn. 1896)).

49. *See, e.g.*, J.A. Kreis & Co. v. City of Knoxville, 237 S.W. 55, 57 (Tenn. 1921) (holding that a party contracting with a municipality is presumed to know the legal limitations placed on municipal officers' authority to bind the municipality).

50. *See, e.g.*, Davis v. Metro. Gov't of Nashville & Davidson County, 620 S.W.2d 532, 535 (Tenn. Ct. App.1981) ("Ignorance of the law is no ground to rescind agreements or to set aside the solemn acts of the parties.").

51. 360 S.W.2d 27 (Tenn. 1962).

52. *See id.* at 29.

53. See *id*

54. White v. Kane, 159 S.W.2d 92, 94 (Tenn. 1942).

55. *See* TENN. CODE ANN. § 32-3-105 (2001); Strong v. Ready, 28 Tenn. 168, 170-71 (1848). In 1783, Massachusetts was the first state to enact an antilapse statute. Erich Tucker Kimbrough, *Lapsing of Testamentary Gifts, Antilapse Statutes, and the Expansion of Uniform Probate Code Antilapse Protection*, 36 WM. & MARY L. REV. 269, 274 (1994).

56. TENN. CODE ANN. § 32-3-105 (2001).

57. *See* Kimbrough, *supra* note 55, at 269.

58. *See* Weiss v. Broadway Nat'l Bank, 322 S.W.2d 427, 432 (Tenn. 1959).

59. *See* TENN. CODE ANN. § 32-3-105(a).

> shall take the estate or interest devised or bequeathed which the devisee or legatee or the member of the class, as the case may be, would have taken, had that person survived the testator, unless a different disposition thereof is made or required by the will.[60]

Tennessee's statute was derived from a provision of the English Wills Act of 1837, also an antilapse statute.[61] Both statutes apply only when the predeceased beneficiary leaves living issue at the time of the testator's death.[62] The Tennessee Supreme Court has determined that the term "issue" refers only to direct descendants of the predeceased beneficiaries.[63]

In 1889, the court reiterated the limited nature of the antilapse statute.[64] In *Dixon v. Cooper*[65], the widow of a predeceased beneficiary attempted to claim a portion of the testator's estate under the antilapse statute.[66] The court clarified the statute's "direct descendants" condition once again by stating: "[T]hese statutes . . . are intended alone to meet the case where a devisee or legatee dies 'leaving issue,' and not every case where such person dies before the testator."[67] The court went on to point out that, under *Strong v. Ready*[68], this was already well-settled law.[69]

---

60. *Id.*

61. *See* Dixon v. Cooper, 12 S.W. 445, 445-46 (Tenn. 1889). The *Dixon* decision states:

> "[W]here any person, being a child or other issue of the testator, to whom any real or personal estate shall be devised or bequeathed, for any estate or interest not determinable at or before the death of such person, shall die in the lifetime of the testator, leaving issue, and any such issue of such person shall be living at the time of the death of the testator, such devise or bequest shall not lapse, but shall take effect as if the death of such person had happened immediately after the death of the testator, unless a contrary intention shall appear by the will.

Id. at 446 (quoting English Wills Act of 1837, 1 Vict., c. 26, § 33 (Eng.)).

62. *See id.* In 1848, six years after the Tennessee antilapse statute was enacted, the Tennessee Supreme Court held that since "the legacy . . . lapsed and remained undisposed of by the will, [the beneficiary] having no issue . . . [the] act of 1842, ch. 177, sec. 3, [has] no application to the case . . . ." Strong v. Ready, 28 Tenn. 168, 170-71 (1848). There is one important difference between the English statute and the Tennessee statute. The English statute only provides for antilapse when the predeceased beneficiary is a "child or other issue of the testator." *Dixon*, 12 S.W. at 445-46. The Tennessee statute is broader. *See* TENN. CODE ANN. § 32-3-105(a). It defeats the lapse of a gift to "any member of a class to which an immediate devise or bequest is made." *Id.*

63. White v. Kane, 159 S.W.2d 92, 95 (Tenn. 1942).

64. *See Dixon*, 12 S.W. at 446.

65. 12 S.W. 445 (Tenn. 1889).

66. *Id.* at 445.

67. *Id.* at 446.

68. 28 Tenn. 168 (1848).

69. *Dixon*, 12 S.W. at 446.

Instead of relying upon the antilapse statute, testators can provide for the disposal of lapsed gifts through a residuary clause contained in the will itself.[70] However, a unique situation arises when a residuary beneficiary predeceases the testator and leaves no issue.[71] As long as there is no other clause directing to whom the lapsed residuary gift should pass upon the death of a residuary beneficiary, the antilapse statute does not apply.[72]

The law has developed two different rules that deal with this situation. The "traditional" rule, which is derived from the English common law,[73] provides that a lapsed gift will pass "to whoever [sic] would have inherited the property in case no will had been made."[74] The Uniform Probate Code's "modern" rule[75] directs that the lapsed residue simply pass to remaining residuary beneficiaries.[76] This rule has become more popular in the United States than the traditional rule.[77]

The Tennessee Supreme Court adopted the traditional rule in *Ford v. Ford*[78] In that case, the testator, Edward M. Ford, executed a will in 1846 which contained specific bequests to his wife and son, Edward C. The will directed that the remainder of Mr. Ford's estate be split between his six other children.[79] In 1847, Mr. Ford executed a codicil disinheriting his daughter Harriet, one of the six children originally named as residuary beneficiaries.[80] Unfortunately, the codicil did not provide for the disposition of her revoked

---

70. *See* Milligan v. Greeneville Coll., 2 S.W.2d 90, 93 (Tenn. 1928).

71. *See* Patricia J. Roberts, *Lapse Statutes: Recurring Construction Problems*, 37 EMORY L.J. 323, 367 (1988) (discussing the current majority view that failed residuary gifts pass to the remaining residuary beneficiaries).

72. *See Milligan*, 2 S.W.2d at 93.

73. Corbett v. Skaggs, 207 P. 819, 820 (Kan. 1922). The traditional rule was "probably laid down by English courts to protect the interests of the primogeniture heir." Jesse Dukeminier, Stanley M. Johnson, James Lindgren & Robert H. Sitkoff, WILLS, TRUSTS, AND ESTATES 387 (Erwin Chemerinsky et al. eds., 7th ed. 2005). This rule is also commonly known as the "no-residue-of-a-residue" rule. *Id.*

74. *Corbett*, 207 P. at 820.

75. UNIF. PROBATE CODE § 2-606 (1975). The Uniform Probate Code project was undertaken in 1962 by the Real Property, Probate and Trust Law Section of the American Bar Association and the National Conference of Commissioners on Uniform State Laws. Robert Whitman, *Revocation and Revival: An Analysis of the 1990 Revision of the Uniform Probate Code and Suggestions for the Future*, 55 ALB. L. REV. 1035, 1042 (1992). Seven years later the first official text emerged. *Id.*

76. UNIF. PROBATE CODE § 2-606(b) (1975). The modern rule provides as follows: "[I]f the residue is devised to two or more persons and the share of one of the residuary devisees fails for any reason, his share passes to the other residuary devisee, or to other residuary devisees in proportion to their interests in the residue." *Id.*

77. *See* Dukeminier et al., *supra* note 73, at 387-88 ("In a majority of states, [the traditional rule] has been overturned by statute or judicial decision . . . .").

78. 31 Tenn. 431, 435 (1852).

79. *Id.* at 432-33.

80. *Id.* at 433.

share of the remainder.[81] Thus, Harriet's former gift lapsed.[82] Therefore, the court had to determine whether this share of the residuary would pass to Edward C., as Edward M.'s intestate heir at law, or to the five other siblings as the remaining residuary beneficiaries.[83]

The court resolved the problem by applying the common law rule of concurrent interests,[84] reasoning that the five remaining residuary beneficiaries had to be either joint tenants or tenants in common.[85] If the beneficiaries were joint tenants, they would be regarded as a single owner of the residuary, with the right of survivorship.[86] In other words, if one beneficiary died, the residuary did not pass to the remaining siblings because, in theory, they already owned it.[87] However, if the beneficiaries were tenants in common, then there was no right to survivorship,[88] the interests would be separate but undivided, and each separate interest could descend to another person.[89] To determine whether the residuary beneficiaries were joint tenants or tenants in common, the court examined the language of the will, which directed that the remainder "be equally divided among them."[90] The court determined that these words expressed the testator's desire for the shares to be distinct and separate.[91] Therefore, the court concluded that the remaining residuary beneficiaries were tenants in common and did not have the right of survivorship with regard to Harriet's revoked share.[92]

Normally, this undistributed share of the estate would fall into the residuary, but in this particular case, the share actually originated in the residuary.[93] Thus, the court introduced the traditional rule to Tennessee case law by stating: "If . . . a bequest to several be of the residue, and it fail as to part by reason of lapse, partial revocation, or other cause before stated, such part cannot fall into the remaining residue, but will be undisposed of and go to the next of kin."[94] The court further reasoned that no evidence suggested a contrary intention on the part of the testator: "He [did] not provide that the other five shall take it, or that the legacies he had given to them should be

---

81. *Id.*
82. *Id.*
83. *Id.* at 435-36
84. *See id.* at 434-35.
85. *Id.* at 435.
86. JESSE DUKEMINIER & JAMES E. KRIER, PROPERTY 340 (Richard A. Epstein et al. eds., 5th ed. 2002).
87. *See id.*
88. *Id.*
89. *Id.*
90. *Ford*, 31 Tenn. at 433.
91. *Id.* at 435.
92. *See id.* at 435-36.
93. *See id.*
94. *Id.* at 435 (citations omitted).

enlarged. It is not reasonable to suppose that, if he had so intended, he would have so provided?"[95]

III. THE "TRADITIONAL" RULE PERSISTS IN TENNESSEE

With its 3-2 decision in *In re Estate of McFarland*, the Tennessee Supreme Court upheld the traditional rule set forth in *Ford* for distributing lapsed residuary gifts.[96] The court adhered to the traditional rule because (1) will construction principles presume the testator's acquaintance with the longstanding *Ford* rule when executing a will[97] and (2) the doctrine of stare decisis requires Tennessee courts to follow the *Ford* decision.[98]

Justice Barker wrote the majority opinion[99] and evaluated whether the lapsed residuary gifts of the McFarland will should pass to the remaining residuary beneficiaries or to the testatrix's intestate successors.[100] At the outset of its analysis, the court explained several principles of will construction.[101] The court stated that the "cardinal rule" of construing a will is to determine the intent of the testator,[102] not to "mere[ly] surmise . . . the testator's intention."[103] Specifically relevant to the *McFarland* case, the court noted that holographic wills garner liberal construction.[104] Next, as urged by the appellant, the court recognized the presumption against intestacy.[105]

Most notably, however, the court presumed that a testator is familiar with pertinent probate law when executing a will.[106] Combining this presumption

---

95. *Id.* at 436.

96. *In re* Estate of McFarland, 167 S.W.3d 299, 306 (Tenn. 2005). *See also Ford*, 31 Tenn. at 435.

97. *McFarland*, 167 S.W.3d at 305.

98. *Id.* at 306. The doctrine of stare decisis has ensured a long life for the traditional rule in the Tennessee. In *Hall v. Skidmore*, the Tennessee Supreme Court held that when overruling a particular decision is "not of the highest value to the community," the rule should stand. Hall v. Skidmore, 171 S.W.2d 274, 276 (Tenn. 1943) (quoting Barton's Lessee v. Shall, 7 Tenn. (Peck) 215, 232 (1823)). In *J.T. Fargason Co. v. Ball*, the Tennessee Supreme Court elaborated. The court noted that the doctrine of stare decisis provides stability and "enable[s] the people to safely judge of their legal rights." *See* J.T. Ferguson Co. v. Ball, 159 S.W. 221, 222 (Tenn. 1913). It also maintained that "radical changes should be made by legislation only." *Id.*

99. *Id.* at 301. Justices Anderson and Birch joined the majority. *Id.*

100. *Id.*

101. *Id.* at 302.

102. *Id.*(citing *In re* Estate of Vincent, 98 S.W.3d 146, 148 (Tenn. 2003).

103. *Id.* (quoting Pinkerton v. Turman, 268 S.W.2d 347, 350 (Tenn. 1954)).

104. *Id.* (citing Garner v. Becton, 212 S.W.2d 890, 891 (Tenn. 1948)).

105. *Id.* at 303 (citing *In re* Walker, 849 S.W.2d 766, 768 (Tenn. 1993)); *see also* TENN. CODE ANN. § 32-3-101 (2001) (requiring that a will be construed as if it had been executed immediately before the testator's death).

106. *McFarland*, 167 S.W.3d at 303 (citing McCarley v. McCarley, 360 S.W.2d 27, 29 (Tenn. 1962)).

with the court's duty to read the will as if it had been signed just before the death of the testatrix, the majority came to its first conclusion supporting the traditional rule by reasoning that "prior to her death," Ms. McFarland must have known that three of her named residuary beneficiaries had died and that she would die partially intestate.[107] The majority concluded its discussion of will construction principles by highlighting a lack of evidence regarding McFarland's wishes or any attempt on her part to remedy the lapse.[108]

Accordingly, the majority proceeded by applying the law governing lapsed gifts.[109] First, the court discussed Tennessee's antilapse statute,[110] stating that it "attempts to further the presumed intent of the testator in the absence of any contrary intent expressed through the will."[111] The court explained that under the statute, lapsed gifts pass to the issue of predeceased beneficiaries.[112] The court found the statute inapplicable in the *McFarland* case, however, because none of the predeceased beneficiaries had left surviving issue.[113] Next, the court noted that instead of relying upon an antilapse statute, a testator may employ a residuary clause to save lapsed gifts.[114] However, the lapsed gifts at issue had originated as part of the residue, and the testatrix had specified no alternative disposition.[115] Therefore, the court concluded that the residuary clause in McFarland's will, like the antilapse statute, could not save the lapsed gifts.[116]

To deal with this predicament, the court explained, the law had developed two options.[117] Under English common law, "the lapsed gift falls out of the terms of the will and passes by intestate succession to the testator's heirs at law."[118] This is the traditional rule adopted by the Tennessee Supreme Court in *Ford*.[119] Under the Uniform Probate Code's "modern" rule, the gift would simply remain in the residue and pass to the remaining residuary beneficiaries in proportion to their interests specified in the will.[120]

The majority acknowledged that the traditional rule had been "widely abandoned" in other states and recognized the reasons for the modern rule's widespread support: "[I]t more closely comports with the probable intent of the

---

107. *Id.*
108. *Id.*
109. *Id.*
110. *Id.* Tennessee's antilapse statute is found at TENN. CODE ANN. § 32-3-105 (2001).
111. *McFarland*, 167 S.W.3d at 303 (citing Weiss v. Broadway Nat'l Bank, 322 S.W.2d 427, 432 (Tenn. 1959)).
112. *Id.*
113. *Id.* (citing Cox v. Sullins, 183 S.W.2d 865, 866 (Tenn 1944)).
114. *Id.* (citing Milligan v. Greeneville Coll., 2 S.W.2d 90, 93 (Tenn. 1928)).
115. *Id.* at 304.
116. *Id.*
117. *See id.*
118. *Id.*
119. *Id. See* Ford v. Ford, 31 Tenn. 431, 436 (1852).
120. *McFarland*, 167 S.W.3d at 304; UNIF. PROBATE CODE § 2-606 (1974).

testator and . . . avoids partial intestacy."[121] Nevertheless, the majority maintained that the traditional rule could not "fairly be termed either incorrect or illogical."[122] Justice Barker wrote, "[I]n our view, it is just as likely that a person would consider the implications of the traditional . . . rule when executing his or her will and thus implicitly intend that lapsed gifts should pass to the heirs, rather than to the remaining residuary beneficiaries."[123]

Next, the majority defended the traditional rule by contending that the rule did not necessarily defy the testator's intent.[124] Again, the court noted that the will did not reveal the testatrix's preferred distribution of the lapsed gifts.[125] The court then reasoned that application of the traditional rule would provide for fifteen of the eighteen residuary beneficiaries exactly as the testatrix instructed in the will, but that the modern rule would adjust the shares of these beneficiaries and "totally frustrate her intentions."[126] Also, the majority pointed out that the traditional rule would pass the lapsed gifts to blood relatives, "the most natural objects of her bounty."[127]

Lastly, the majority defended the traditional rule under the doctrine of stare decisis.[128] The appellant argued that the testatrix's probable intent and the widespread adoption of the modern rule favored abandonment of the traditional rule.[129] However, the court did not find these arguments persuasive and stated that "[r]adical changes in the law are best made by the legislature."[130]

In the dissenting opinion, Chief Justice Drowota advocated the adoption of the Uniform Probate Code's modern rule for three reasons.[131] First, Chief Justice Drowota stated his belief that the modern rule would more likely reflect the actual intent of a testator.[132] He admitted, however, that both the traditional and modern rules "involve a certain degree of arbitrariness . . . ."[133] Even so, Drowota argued that the testator's choice to include a residuary clause evidences a preference for the remaining residuary beneficiaries over intestate heirs.[134] The dissent believed that increasing the remaining beneficiaries' shares of the residuary would be more consistent with the testatrix's intent than passing the lapsed gifts to people who may not have been named in the will.[135]

---

121. *McFarland*, 167 S.W.3d at 304 (citations omitted).
122. *Id.*
123. *Id.* at 305.
124. *See id.*
125. *Id.*
126. *Id.*
127. *Id.* (citing Furchtgott v. Young, 487 S.W.2d 301, 304 (Tenn. 1972)).
128. *See id.* at 305-06 (citations omitted).
129. *Id.* at 305.
130. *Id.* at 306.
131. *See id.* at 306-08 (Drowota, C.J., dissenting).
132. *Id.* at 307 (citing Corbett v. Skaggs, 207 P. 819, 821 (Kan. 1922); In re Frolich's Estate, 295 A.2d 448, 451-52 (N.H. 1972); In re Gray's Estate, 23 A. 205, 206 (Pa. 1892)).
133. *Id.*
134. *Id*; *see also infra* note 156.
135. *Id.*

Next, Drowota noted that the "purpose of a residuary clause is to function as a dragnet for devising parts of [an] estate not otherwise disposed of."[136] Given this purpose, the dissent found no fault with gathering the "lapsed residuary gifts" and distributing them to "surviving residuary beneficiaries."[137]

Second, Drowota argued that the modern rule better complements the presumption against intestacy "which is itself aimed at reinforcing testamentary intent."[138] The dissent quoted *Williamson v. Brownlow*[139] which stated: "[I]t will be presumed that one who undertakes to make a will does not intend to die intestate as to any of his property."[140] The dissent further asserted that the residuary clause itself was evidence that the testatrix did not want any part of her estate to pass by intestate succession.[141]

Third, the dissent argued that the adoption of the modern rule would not offend the doctrine of stare decisis.[142] Drowota pointed out that the Tennessee Supreme Court had never reaffirmed the *Ford* decision and that the Tennessee Court of Appeals had only applied it once.[143] In conclusion, the dissent quoted the following passage from *Metropolitan Government of Nashville & Davidson County v. Poe*:[144] "[The court] should not close its doors to the changing conditions, but should so fashion its opinions that the new truly grows out of the old as the product of a changing environment."[145]

IV. THE TESTATOR'S INTENT V. STARE DECISIS D/B/A THE PRESUMPTION OF ACQUAINTANCE WITH THE LAW

In *In re Estate of McFarland*, the Tennessee Supreme Court reaffirmed the 150-year-old *Ford* rule by relying on a rule of will construction that presumes the testator's acquaintance with the law.[146] The majority essentially equated this presumed knowledge of the law with an intent to die partially intestate.[147] The majority believed that a testator would just as likely anticipate the operation of the traditional rule as she would prefer the lapsed gift to pass to individuals already named in the will.[148] Although this may be the case, it seems less likely when the will in question is a holographic will, which, under

136. *Id.* (citing Oliver v. Wells, 173 N.E. 676, 678 (N.Y. 1930)).

137. *Id.* (citing *Corbett*, 207 P. at 822).

138. *Id.* (citations omitted).

139. 410 S.W.2d 878 (Tenn. 1967).

140. *McFarland*, 167 S.W.3d at 307 (Drowota, C.J., dissenting) (quoting Williamson v. Brownlow, 410 S.W.2d 878, 880-81 (Tenn. 1967)).

141. *Id.*

142. *Id.* at 308.

143. *Id.* (citing Davis v. Anthony, 384 S.W.2d 60, 63 (Tenn. Ct. App. 1964)).

144. 383 S.W.2d 265 (Tenn. 1964).

145. *McFarland*, 167 S.W.3d at 308 (Drowota, C.J., dissenting) (quoting *Poe*, 383 S.W.2d at 277).

146. *See id.* at 303 (majority opinion).

147. *See id.*

148. *Id.* at 305.

Tennessee law, must be liberally construed.[149] A layperson who drafts a holographic will may not anticipate the effects of the traditional rule. It is more likely that a layperson will believe that only persons named in the will can receive shares of the estate. The dissent is more realistic in its stance on Ms. McFarland's intent.[150] She almost certainly possessed an unsophisticated understanding of the law.

The majority correctly noted that there was a lack of conclusive evidence that Ms. McFarland did not intend to die partially intestate.[151] However, the available evidence actually suggested the opposite.[152] The mere presence of a will is evidence of the testator's intent to avoid intestacy.[153] Here, Ms. McFarland executed a will that contemplated the complete disposal of her estate.[154] She even included a residuary clause.[155] The dissent correctly notes that a residuary clause only reinforces the presumption against intestacy.[156] In fact, Ms. McFarland's will very nearly defeated intestacy. If the predeceased beneficiaries had simply left surviving issue, the antilapse statute would have provided for the estate's full disbursement.[157] The majority brushed aside the presumption against intestacy by assuming that Ms. McFarland's presumed acquaintance with the law manifested her intent to die intestate.[158]

The majority maintained that adopting the UPC's modern rule would "completely redraft the terms of the will."[159] It is true that the percentage assigned to each of the fifteen remaining residuary beneficiaries would increase slightly under the modern rule.[160] In addition, the majority noted that "[McFarland] clearly gave considerable thought to the specific percentages of her estate which she allotted to each beneficiary."[161] It follows that she also must have given considerable thought to the persons named in the will and not to those persons excluded. For example, Ms. McFarland did not provide for

---

149. *See* Garner v. Becton, 212 S.W.2d 890, 891 (Tenn. 1948).

150. *See McFarland*, 167 S.W.3d at 306 (Drowota, C.J., dissenting).

151. *Id.* at 303 (majority opinion).

152. As the administrator argued, the inclusion of eighteen residuary beneficiaries suggested that McFarland did not intend for any portion of her estate to pass by intestate succession. *See id.* at 301-303.

153. *See id.* at 303 (citing *In re* Walker, 849 S.W.2d 766, 768 (Tenn. 1993)); *see also* TENN. CODE ANN. § 32-3-101 (2001).

154. *See McFarland*, 167 S.W.3d at 301.

155. *Id.*

156. *Id.* at 307 (Drowota, C.J., dissenting) (citing *In re* Frolich's Estate, 295 A.2d 448, 452 (N.H. 1972)). "The rationale for the current majority view is that it is consistent with intent and consistent with the notion that the purpose of the residuary clause is to prevent intestacy." Patricia J. Roberts, *Lapse Statutes: Recurring Construction Problems*, 37 EMORY L.J. 323, 368 (1988).

157. *See McFarland*, 167 S.W.3d at 302.

158. *See id.* at 302-03.

159. *Id.* at 305.

160. *See id.*

161. *Id.*

her intestate heir, James Cox.[162] If the majority had adopted the modern rule, only individuals named in the will would have received shares of the estate.[163]

Also, because the beneficiaries took percentages of the estate under the residuary clause, Ms. McFarland did not include a fixed dollar amount.[164] Thus, it must have been more important to her that some beneficiaries receive more of her estate than it was for each beneficiary to receive a fixed sum.[165] If the court had adopted the modern rule, this wish would have been fulfilled; moreover, an unnamed heir would not have taken thirty percent of her estate.[166] The dissent was more logical in its discussion of this matter. Because the testator intends for a residuary clause to operate as a "dragnet" for remainder shares of the estate, the dissent thought it "entirely sensible to gather up lapsed residuary gifts and fold them back into the shares of surviving residuary beneficiaries."[167]

The majority believed that adopting the "modern" rule would have constituted a "radical change in the law."[168] In reality, the modern rule would only affect a small number of cases, namely lapsed residuary disputes. In the 153-year span between the *Ford* and *McFarland* decisions, the traditional rule was only applied by a Tennessee appellate court once.[169] Because an overwhelming majority of states have adopted the modern rule, it is clear that public policy in the area of will construction has dramatically shifted since the 1852 *Ford* decision.[170]. The scope of *McFarland* is not limited to the facts of that particular case. The holding demonstrates an adherence to old-fashioned, traditional will interpretation principles, which will dictate the outcome of every subsequent Tennessee case involving a lapsed residuary gift.[171] Although

---

162. *See id.* at 301; *see also supra* note 20.

163. *See McFarland*, 167 S.W.3d at 304-05.

164. *See id.* at 301.

165. *See id.*

166. *See id.* at 304-05. *See also supra* note 12.

167. *McFarland*, 167 S.W.3d at 307 (Drowota, C.J., dissenting) (citing Corbett v. Skaggs, 207 P. 819, 822 (Kan. 1922); Oliver v. Wells, 173 N.E. 676, 678 (N.Y. 1930)).

168. *Id.* at 306 (majority opinion).

169. *Id.* at 308 (Drowota, C.J., dissenting) (citing Davis v. Anthony, 384 S.W.2d 60, 63 (Tenn. 1964)).

170. *See id.* at 304 n.5 (majority opinion)..

171. John Langbein states:

> The puzzle about the Wills Act formalities is not why we have them, but why we enforce them so stringently. Consider *Groffman* . . . [E]ach of the two witnesses . . . took his turn signing the will in the dining room while the other witness was in the living room. The will was held invalid for violation of the requirement that the testator sign or acknowledge it in the presence of two witnesses present at the same time, although the judge forthrightly declared: 'I am perfectly satisfied that that document was intended by the deceased to be executed as his will and that its contents represent his testamentary intentions.' The Wills Act is meant to implement the decedent's intent; the paradox in a case like *Groffman* is that the Wills Act defeats that intent.

John H. Langbein, *Excusing Harmless Errors in the Execution of Wills: A Report on*

the decision will affect only a small number of cases, the consequences are dramatic. And because the number of affected cases is so small, it is doubtful that the legislature has much incentive to remedy the situation. Furthermore, the decision encourages continued adherence to strict will interpretation principles among the lower courts.

Just as the Tennessee Supreme Court declined to overrule *Ford*, it could have saved the legislature the effort of enacting a new statute by adopting the modern rule. Perhaps, rather than leaving "radical changes" to the General Assembly, the court actually intended to discourage holographic wills. After all, holographic wills are valid in only half of the states and "have consistently spawned litigation"[172] for reasons similar to the circumstances in *McFarland*.

In Tennessee, holographic wills are authorized by statute.[173] However, the *McFarland* court seemed to discourage the use of holographic wills by moving away from liberally construing them.[174] In this case, the majority places the presumption of acquaintance with the law above the presumption against intestacy. This diminishes the cardinal rule of will construction—ascertaining the testator's intent, especially when the testator is uneducated in the law. As the dissent pointed out, the "presumption against partial intestacy . . . is itself aimed at reinforcing testamentary intent."[175] Although the majority stated that courts must avoid "mere surmise as to the testator's intention," it overlooked several aspects of the "context, general scope, and purpose of [Ms. McFarland's will]."[176] Because holographic wills are valid in Tennessee, the courts should afford them the liberal construction they are due.[177]

V. CONCLUSION

In *McFarland*, the Tennessee Supreme Court upheld an 1852 decision that left lapsed residuary gifts to intestate heirs rather than remaining residuary beneficiaries. In reaching this decision, the court applied traditional will construction principles and adhered to the doctrine of stare decisis. Unfortunately, the *McFarland* decision was more technically sound than it was

---

Australia's Tranquil Revolution in Probate Law, 87 COLUM. L. REV. 1, 3-4 (1987) (citations omitted).

172. Kevin R. Natale, Note, *A Survey, Analysis, and Evaluation of Holographic Will Statutes*, 17 HOFSTRA L. REV. 159, 160-61 (1988).

173. *See* TENN. CODE ANN. § 32-1-105 (2001).

174. *See McFarland*, 167 S.W.3d at 302.

175. *Id.* at 307.

176. *See id.* at 302.

177. *See* TENN. CODE. ANN. § 32-1-105 (2001).

> The main concern is in reducing the number of occasions in which holographic wills are invalidated on a technical basis, not in simplifying a court's task or eliminating litigation. If these latter concerns are paramount to a legislature, then holographic wills should not be authorized at all. Indeed, it is the premise of this Note that there should be either liberal provision for holographic wills or no provision at all.

Natale, *supra* note 172, at 201.

logical. Therefore, Tennessee remains stuck with an outdated, impractical rule that the Tennessee Supreme Court had difficulty defending. In defaulting to precedent, the court seems to have washed its hands clean of the issue. It is now up to the legislature to permanently place the *Ford* rule on its dusty library shelf and acquaint Tennessee probate law with modern times.

JULIE M. COCHRAN

TENNESSEE LAW REVIEW

ANNUAL INDEX
VOLUME 73
2005-2006

THE UNIVERSITY OF TENNESSEE
COLLEGE OF LAW

INDEX TO VOLUME 73

Tennesee Law Review Association, Inc.
1505 West Cumberland Avenue
Knoxville, Tennessee 37996-1810

AUTHOR INDEX

SUBJECT INDEX

TABLE OF CASES

A

D

E

F

G

H

I

J

M

N

Q

R

S

T

U

V

Y

Z